Essays on Trade, Human Capital and Industry Structure

Benedikt Dengler

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 11 March 2019

European University Institute
Department of Economics

Essays on Trade, Human Capital and Industry Structure

Benedikt Dengler

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Piero Gottardi, EUI and University of Essex (supervisor)
Prof. Arpad Abraham, EUI
Prof. Simon Evenett, University of St. Gallen
Prof. Enghin Atalay, University of Wisconsin-Madison

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I, Benedikt Hermann Dengler, certify that I am the author of the work *Essays on Trade, Human Capital and Industry Structure* I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

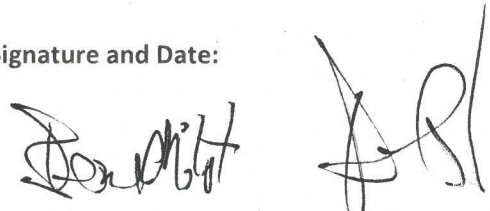
Statement of inclusion of other work:

I confirm that chapter 1 was jointly co-authored with Vinzenz Ziesemer and I contributed 50 % of the work.

I confirm that chapter 2 was jointly co-authored with Bernard Hoekman and I contributed 60 % of the work. The chapter has previously been published as CEPR Discussion Paper No. 13266.

I confirm that chapter 4 was jointly co-authored with Árpád Ábrahám and T. Kirk White and I contributed 60 % of the work.

Signature and Date:



Benedikt Dengler, February 24, 2019

Abstract

This thesis consists of four chapters, covering a range of economic fields and methodologies.

Chapter 1, joint work with Vinzenz Ziesemer, investigates how skill supply incentives and general equilibrium wage effects shape the earnings distribution, connecting two separate strands of the literature on earnings inequality. On the one hand, the literature on skill-biased technological change studies how general equilibrium effects between different types of workers shape relative earnings and thus the income distribution. On the other hand, the literature on taxation suggests that *incentives* to accumulate human capital drive the earnings distribution. Combining data on occupational skills and US Census data, we first show empirically that job skills can be summarized by two dimensions: manual and cognitive skills. We show further that the latter strongly correlates with traditional measures of schooling, while the former does not. Based on these observations, we build a model featuring both human captial accumulation incentives for cognitive skills and general equilibrium wage effects between manual and cognitive skill prices. We theoretically investigate the response of the earnings distribution to changes in the tax progressivity which alter the incentives to accumulate cognitive human capital. It is shown that changes in tax progressivity like those occurring during the second half of the 20th century can lead to polarization in the labor market. A calibrated version of the model predicts these effects to be relatively small quantitatively.

Chapters 2 and 3 empirically investigate the effects of international agreements on import shares in the government sector, focusing on different types of agreements and methodologies. Chapter 2, joint work with Bernard Hoekman, employs a cross-country panel regression framework to study the effect of the WTO Government Procurement Agreement (GPA) on government sector import shares in the aftermath of the 2008 financial crisis. A growing trade policy literature investigates the potential role of international agreements in reducing future policy uncertainty and thereby facilitating trade. This potentially provides an alternative channel for the GPA to impact trade in addition to faciliating reciprocal market access. Our results are suggestive of such international disciplines acting as an effective commitment device: GPA membership is associated with

a significantly higher import share following the 2008 financial crisis than is observed for countries that are not members. In addition, there is evidence that the GPA and PTAs that cover public procurement are partial substitutes.

Chapter 3 focuses on the relationship between the depth of PTA public procurement provisions and their trade effects. Recently, an increasing number of PTAs include public procurement chapters, characterized by an increasing level of detail. The emergence and proliferation of detailed PP provisions in PTAs raises the question of whether these more substantive types of agreements have a real effects on bilateral trade in a systematic way. In this chapter I study to what extent the proliferation and deepening of PP provisions in recent PTAs leads to tangible changes in the tradeflows they cover, using a bilateral gravity framework. Previous research has not found real effects of PP openness commitments and attributed this to a lack of legal enforceability (Rickard and Kono 2014). I revisit this question using more direct measures of PP imports and a novel classification of PTA PP provisions by Shingal and Ereshchenko (2018), explicitly designed to distinguish PP provisions by their legal enforceability. I find that the presence of deep PP provisions is indeed associated with significantly higher trade volume absorbed by the government sector in the importing country. When looking at government imports disaggregated by economic sector, it becomes clear that this aggregate finding is driven primarily by service sector imports. Furthermore, the real impact of PP provisions seems to be driven primarily by the European Union. The results are not robust to excluding intra-EU bilateral flows.

Chapter 4 is joint work with Árpád Ábrahám and T. Kirk White. The chapter describes a research proposal aiming to study the effect of the ownership structures along supply chains on firms' bargaining power in input markets. Two recent contributions document that (a) supply chain considerations are not the prime focus of vertical integration and (b) input prices display substantial heterogeneity across U.S. firms. In this paper, we propose to empirically test whether vertical integration and input price dispersion are related, using unique features of U.S. Economic Census micro data that have already been employed by Atalay et al. (2014) and Atalay (2014). In particular, we argue that owning productive capacities upstream endows firms with informational or bargaining advantages which result in lower procurement prices for their downstream units. We propose to decompose input price dispersion in a between-supplier and a within-supplier component and investigate how both components contribute to the input price advantage of vertically integrated firms. This project sheds light on the determinants of measured firm productivity. Input price dispersion through the between-supplier component predicts measured productivity dispersion through supplier selection, while within-supplier price dispersion implies that measured productivity dispersion is driven by differences in supplier market power.

Acknowledgments

First and foremost, I would like to thank my advisors, Piero Gottardi and Árpád Ábrahám, for their guidance and support over the past years. Both have always been available, also on short notice, to discuss any issue I might have had and I learned a lot from them. Enghin Atalay and Simon Evenett kindly agreed to become part of the committee of this thesis. I would like to thank them for taking the time to read and comment on my work. I'd also like to thank Philipp Kircher and especially Bernard Hoekman. Working for and with both of them has been integral for the completion of this thesis.

Three out of the four chapters of this thesis are co-authored. First, I'd like to thank Vinzenz. Not only did we work together on several projects – some more successful than others – but also for all my other academic endeavors he was always up for a discussion. I'd also like to thank Árpád for putting me in touch with Kirk and Kirk for his availability and patience in our coordination-intensive trans-atlantic collaboration. Finally, I'd like to thank Bernard again. He gave me a new perspective on research when I was in need and two chapters in the present thesis grew out of my work with him.

The EUI is an exceptional place to write a PhD. I would like to thank the German Academic Exchange Service (DAAD) and the EUI for enabling me to be part of this community over the past years.

During my time in Florence, I am glad to have met a lot of people without whom the past years would not have been the same. Thanks to David and Simon, João and Fabian, Vinzenz, Agnieszka, Liiri and Francesca, Ginevra, Noa, Daniel, Gözde, Ada, Tobias and Mariana. Thanks to my entire econ-cohort.

I would like to thank my family, my parents and my siblings Josefa, Theresia and Kilian. I've always had their support and I consider myself very lucky to have them in my life.

This thesis is dedicated to the memory of my grandfathers, Hermann Dengler and Alfred Fettweis. Each in their own way, both have been shaping my thinking to this day.

Contents

Chapter 1

Polarization: A Supply-Side Mechanism

This Chapter is joint work with Vinzenz J. Ziesemer.

Abstract

The literature on earnings inequality consists of two disparate strands. On the one hand, skill-biased technological change describes how general equilibrium effects between different types of workers shape the income distribution. On the other, the literature on taxation suggests that incentives to accumulate human capital drive the earnings distribution. This paper combines both approaches, underpinned by an empirical analysis of occupational skill data. We find that incentive changes in taxation like those that occurred in the 2$^{\text{nd}}$ half of the 20$^{\text{th}}$ century can lead to polarization of the labor market. This supply- or policy-driven explanation is complementary to the demand- or technology-driven explanations in the existing literature.

1.1 Introduction

1.1.1 Motivation

The effect of tax incentives on the formation of human capital have been studied extensively, both in positive (e.g. Guvenen, Kuruscu, and Ozkan, 2014) and normative (e.g. Stantcheva, 2017) contexts. This paper studies the incentive effects of taxation when human capital is multi-dimensional and the labor market is cleared in general equilibrium. Our results shed new light on relative movements of the earnings distribution. First, the presence of multi-dimensional skills can rationalize the relative unresponsiveness of low earnings to

tax incentives. Second, we present a novel labor supply-driven mechanism for polarization of the earnings distribution to arise. The interaction between accumulation incentives and general equilibrium effects turns out to be key for the generation of non-monotone changes to the earnings distribution.

Tax progressivity is substantially different both across countries and over time. Since the 1970s, tax levels and progressivity in the United States have fallen dramatically. Guvenen, Kuruscu, and Ozkan (2014) estimate tax schedules on OECD data for the years 1983 and 2003. Average tax rates for the US are depicted in Figure 1.1. [1] Guvenen, Kuruscu, and Ozkan (2014) argue that high levels of taxes, and especially tax progressivity, play an important role in shaping the earnings distribution by reducing optimal human capital investments, particularly for the highly skilled. Taxes then compress the distribution of earnings. Guvenen, Kuruscu, and Ozkan argue that their mechanism explains changes in earnings inequality both across countries and over time, showing that it explains up to two-thirds of the change in the US wage premium between 1973 and 2003.[2]

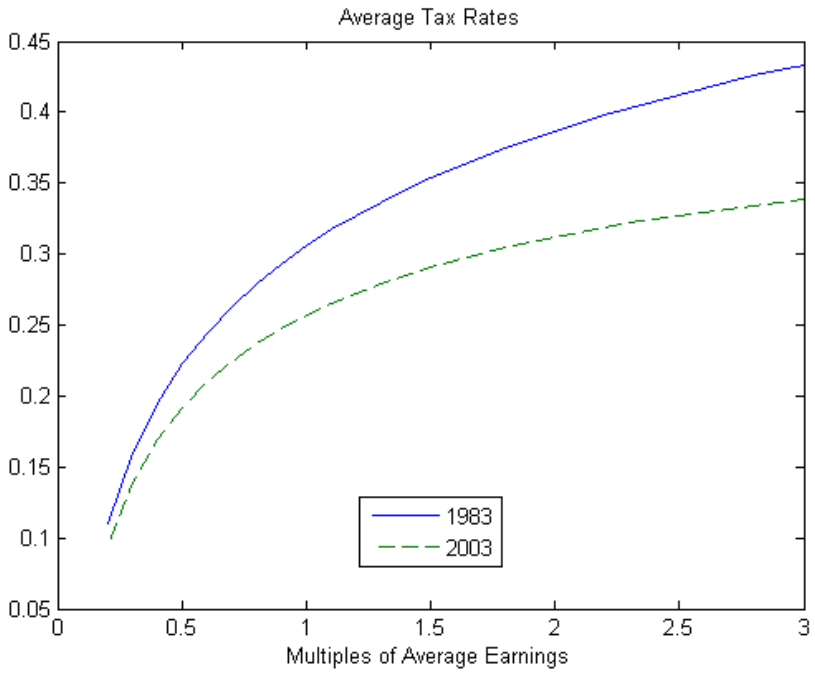
While Guvenen, Kuruscu, and Ozkan (2014)'s results are suggestive, Skill-Biased Technological Change (SBTC) has been the main explanatory model of why inequality grew so strongly over recent decades in the United States (Katz and Murphy, 1992) and elsewhere (Berman, Bound, and Machin, 1998). According to the SBTC theory, increased supplies of highly educated labor keep the wage premium to education down, while technological change that is biased towards skilled labor continuously drives it up. In the 1980s the growth of educational attainment slowed down, explaining why inequality took off. In this type of theory, human capital is essentially considered two-dimensional, and general equilibrium effects are the main driver.

The perhaps primary challenge to both types of models is the 'polarization' phenomenon: the observation that starting in the 1980s jobs in the middle of the earnings distribution have seen less growth in wages and employment than those at the top or bottom, both in the US (Autor and Dorn, 2013) and across advanced economies (Goos, Manning, and Salomons, 2014). This coincided with growth in overall earnings inequality, i.e. a growing

[1] Figure 1.6 in Section 3.2 plots cross-country differences in tax progressivity and the relationship to inequality measures.

[2] All this is much in line with ideas from a public economics literature that considers how taxes are set optimally when human capital is endogenous (see Bovenberg and Jacobs (2005) for a static setting, and Stantcheva (2017) for a dynamic extension). In this line of research, human capital is considered one-dimensional, and general equilibrium effects on wages play no role. Recently, a literature has developed that considers the original Mirrlees problem when many types interact in general equilibrium. A recent contribution is by Sachs, Tsyvinski, and Werquin (2016). The formation of types (or human capital) has so far been taken as exogenous. The same applies to previous work by Teulings (2005), which provided a framework for tracing out general equilibrium effects across many types.

Figure 1.1: Average tax rates in the United States



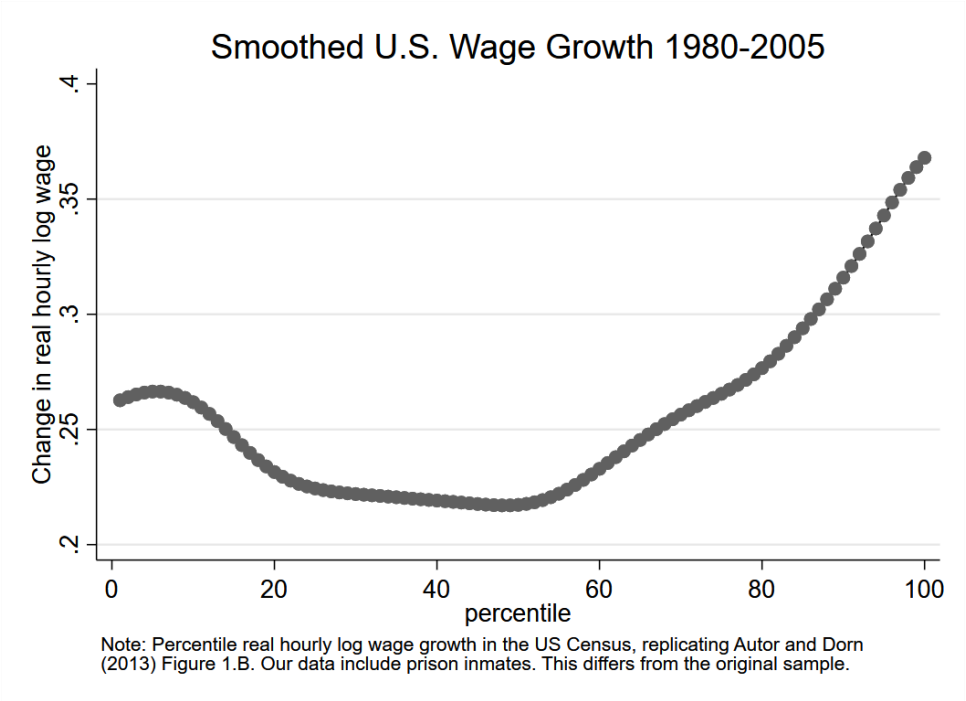
difference between the top and the bottom. Figure 1.2 displays these phenomena for the United States. A number of explanations have been put forward to explain what is different about jobs in the middle of the income distribution, such as offshorability or competition from China or declines in unionization rates in the manufacturing sector, but consensus has formed around the view that these jobs have a higher degree of 'routineness', and are therefore more susceptible to automation (by machines, robots, and computers). In short, polarization of the labor market is seen as demand-driven, and attributed to exogenous technological forces. See Autor et al. (2010) for a review of this literature.

This paper takes an entirely different and complementary approach. It extends the analysis of incentive changes as in Guvenen, Kuruscu, and Ozkan (2014) to a multi-dimensional setting, in which there is also a role for the general equilibrium effects of Katz and Murphy (1992). Thus, it combines both of the established approaches to earnings inequality discussed above.[3]

We begin our analysis with a set of empirical results. Analyzing data on occupational skills from the Dictionary of Occupational Titles (DOT) combined with Census data, this paper establishes that there seem to be *two* relevant dimensions of job skills: cognitive and manual skills, both of which can be thought of as continuous variables. We find the

[3]A similar setup has been used by Guvenen and Kuruscu (2010) and Guvenen and Kuruscu (2012) to study skill-biased technological change. However, in both papers the general equilibrium wage effect is deliberately shut off by choosing a linear production technology. This precludes interaction effects between relative skill quantities and prices of the type we study in this paper.

Figure 1.2: Wage Inequality Growth and Polarization in the United States



importance of these skills to be heterogeneous over the distribution of earnings: manual skills play a relatively larger role at the bottom of the distribution, cognitive skills play a larger role at the top. Similar to Katz and Murphy, the cognitive skills coincide heavily with schooling decisions. We discuss the details of this analysis in Section 1.2.

Taken together, these points have important implications for the impact of tax incentives: First, those incentives are more relevant for those at the top of the income distribution than those at the bottom. This is because cognitive skills are subject to individual investment – and therefore incentives – to a much larger extend than manual skills, and because cognitive skills dominate at the top of the income distribution. Second, in general equilibrium, a change in the relative amount of cognitive skills may affect the relative prices of the two skill types and therefore individual earnings - a channel that is absent in models of one-dimensional human capital, but common in the literature on SBTC.

Motivated by these empirical findings, Section 1.3 sets out a simple life-cycle model in which earnings are derived from cognitive and manual skills, cognitive skills are subject to endogenous investment decisions and relative wages are determined in general equilibrium. Importantly, in our model of skills it is not different education levels that map into different skill types as is standard in the literature.[4] Instead, both skill types

[4]One notable exception is Lindenlaub (2017). She analyzes a matching model with multi-dimensional skills and shows how different rates of technological change between different skills can lead to polarization. Unlike in our model, she takes all skills as exogenously given.

are continuous, and one of them is formed by education. In Section 1.4 we use a simpler and more tractable representation of the model to study the effects of tax progressivity in our setting theoretically. We emphasize two implications: inequality and polarization.

First, our mechanism impacts *inequality*. Just like in the uniform human capital model, a decline in tax progressivity impacts the top of the income distribution more than the bottom, thereby increasing income inequality in absolute terms. This is because, in typical human capital models, the more able spend more time learning, and tax progressivity reduces the private gains from having learned earlier in life.

Two things are different in the multi-dimensional case, causing *polarization* to arise. First, the heterogeneous impact of taxes becomes much stronger, so that income inequality also increases in relative terms. This is because lower tax progressivity increases the relative supply of cognitive skills more than that of manual skills, thereby increasing the latter's relative price. This lowers incentives to acquire cognitive skills. Second, the increasing relative price of manual skills, which are more important at the bottom of the income distribution, makes earnings at the bottom of the distribution even less sensitive to progressivity. If this effect is strong enough, it can even increase the wages of those at the bottom relatively more than of those at middling levels of the distribution. In short, the tails of the distribution potentially respond relatively stronger to changes in tax progressivity, with a reduction in progressivity causing polarization in earnings. We discuss the underlying mechanism in further detail in Section 1.4.

The earnings distribution is likely subject to a multitude of economic forces and no single mechanism will be able to fully account for the changes that took place during the period in which polarization arose. We argue that our mechanism potentially contributes in addition to the existing mechanisms put forward in the literature, but do not claim it as the sole driver of the observed changes. In this paper, we attempt to study our supply-side mechanism in isolation. That limits the extent to which we can observe its empirical implications in data. Nevertheless, we include a qualitative comparison of the model's macro-economic implications to data, both over countries and across time, in Section 3.2. The model predicts non-linear patterns of (relative) change in income distributions due to tax changes. While, as we discuss in detail, some of these patterns could also be generated by alternative setups with one-dimensional human capital, our theoretical exposition demonstrates that multi-dimensional models have a much easier time to account for such movements. Using OECD data on income distributions across countries and over time, we construct measures of income inequality and measures of tax progressivity across countries. Patterns that are easily accounted for by our multi-dimensional model are pervasive. We also discuss implications for changes over time, and the limitations present in verifying these.

Section 1.6 takes an enriched version of the model developed in Section 1.3 that can be solved numerically. Parameters are now tied down so that the baseline version of the enriched model matches relevant moments of the US economy. The model is then used to study the quantitative impact of a typical decline in tax progressivity.[5] In order to do so, we use the decline of US tax progressivity since the 1980s that we discussed above as our experiment. We calibrate the model economy to match moments from the US economy in the early 2000s. We then compare the steady-state earnings distribution of this economy with the counter-factual tax progressivity of 1983 to the one in 2003 and calculate the rate of change in earnings.

The main goal of this exercise is to gauge the general quantitative "bite" of the human capital investment channel on changes in the earnings distribution, rather than wanting to account for the empirical change in earnings growth over the same period. This would require at the very least taking into account the transition period as well as cohort composition effects, both of which our model is silent on. More generally, we are looking at our mechanism in isolation, whereas in reality several channels are likely to have played a role in the rise of polarization. The results from the experiment indicate that the model captures growth in overall wage inequality reasonably well. Most of the change comes out of the upper half of the income distribution, in line with the empirical evidence. The results further indicate that the polarization effect exists, and is quantitatively sizable but smaller than what we observe empirically. In conclusion, our mechanism has impact under quantitatively relevant variations in policy. We also argue why our estimate of the mechanisms quantitative implications might be seen as a lower bound.

Our results contribute to two separate literatures. First, they provide a natural explanation for the lack of response in the lower half of the income distribution to changes in the human capital accumulation incentive structure. Existing papers focusing on uniform human capital, such as Guvenen et al. (2014), lack explanatory power in this region of the distribution. By adding the general equilibrium relative price effect, our model complements the direct incentive effect studied in their paper with the general equilibrium price effect. The latter works primarily in the lower half of the distribution and helps to limit the increase in total inequality.

Second, existing theories of polarization are primarily labor demand driven. Autor and Dorn (2013) introduce a third 'routine' skill category and explain polarization through

[5]In this paper, we focus on tax progressivity. As Guvenen, Kuruscu, and Ozkan demonstrate, allowing for flexible labor supply makes tax levels a disincentive in the accumulation of human capital as well.In the context of a cross-country comparison, Guvenen, Kuruscu, and Ozkan find that differences in tax progressivity are a more important determinant of differences in inequality than are differences average tax rates, explaining our focus on the former.

increasing automation of 'routine'-intensive tasks, reducing the demand for jobs located in the middle of the skill distribution. By adding general equilibrium relative price effects to the traditional skilled-unskilled dichotomy of the endogenous human capital literature, we are able to generate qualitatively similar changes in the earnings distribution without resorting to a third type of skill. Given the complexity of the earnings distribution, there are likely many underlying factors at work. Consequently, we see our channel as complementary to the skill-demand based explanations put forward in the existing literature.

Throughout, we emphasize that taxation is just one particular type of disincentive to human capital formation. In principle, there are many other distortions driving a wedge between public and private returns to education that differ across parts of the population. Two of the major trends in the 2nd half of the 20th century have been declines in gender and race based discrimination, both in education and the labor market. Hsieh et al. (2016) attribute about 25% of total economic growth in the US between 1960 and 2010 to changes in discrimination against women and minorities. It also seems reasonable to think that these trends correspond to a reduction in wedge progressivity, since discrimination seems like a particularly salient issue towards the top of the distribution. Thus, for the remainder of this paper, one may want to think of 'wedges' more generally whenever we discuss taxes. Our quantitative results regard taxation only, so that investigating the role of discrimination for polarization is our main suggestion for further research. Section 1.7 concludes and provides further such directions for future research.

1.2 Manual and Cognitive Skills

We use data on the skill content of a number of occupations provided by the Dictionary of Occupational Titles. We analyze the structure of these data using a statistical technique (Principal Compenent Analysis) that allow us to reduce the dimensionality of the data and subsequently interpret them. We fin that skills are best summed up by two dimensions: cognitive and manual skills, both of which are important. In order to map the skill content into the wage distribution, we link the DOT data to the Census. This also enables us to investigate how skill measures relate to other observables, such as education. It turns out that the cognitive component of skills is strongly correlated to measures of education, while the manual component is not. Below, we discuss our data sources in more detail. Empirical methodology and results are presented thereafter.

The main drawback of our type of analysis is that we look for underlying skill categories in the data *per se*, i.e. not in relation to the wages or schooling decisions we expect them to explain. Our results in first instance only aim to have explanatory power with regards to the questions and answers observed in the DOT. Several arguments speak for our

approach nevertheless. First, the clear advantage of this approach is that our measures are in some sense still direct measures of skills, even if they are compounded and rely on analysts. Any explanatory power they have in our further analysis is not due to how we have produced them. Second, and related, the questions included in the DOT were included for a reason: because they were believed to be relevant measures of occupational skill. Last but not least, there is a related literature in which data on skills are directly related to wages. The most important reference in this regard is Yamaguchi (2012), who estimates a structural model of wage development in relation to unobserved skills using the same data on occupational skills as we do. He also finds two underlying skill factors to be of major importance, which he refers to as cognitive and motor tasks.

1.2.1 Data

DOT

We use the 'Current Population Survey (CPS), April 1971, Augmented With DOT Characteristics and Dictionary of Occupational Titles (DOT)', obtained from the IPCSR. This version of the CPS was augmented with data on occupation characteristics from the 4th edition of the Dictionary of Occupational Titles (DOT). The 4th edition of the DOT is unique, in the sense that it is the final edition of the so-called 'Analyst Database'. Over centuries, starting in the mid-1930s, the United States Employment Service led an effort to systematically document the skills required to perform a range of occupations. This was done by sending trained occupational analysts to job sites, where they would complete standardized questionnaires on occupation content. While the database was revised since, the focus after the 4th edition of the DOT shifted to the generation of O*NET data, which are based on surveys of employees and employers, and therefore much less suitable for comparison across occupations. Following much of the literature, we therefore choose to use the 4th edition DOT. The main advantage of using the augmented CPS database is that it provides us with numbers of workers per occupation in the original DOT occupation classification.

The 4th edition DOT provides information on 46 variables of skills needed for and characteristics of 3886 DOT occupations (some examples: Marine Architect, Die-Designer Apprentice, Weather Observer, Hypnotherapist). In the nationally representative CPS database that we use, we also have a proportion of the working population for each of these occupations. The 46 variables consist of the analyst's answers to a wide variety of questions per occupation:

1. To what extent does the job relate to data, people, things? (3 questions)

2. What educational development is required (reasoning, mathematical, language, vocational)? (4 questions)

3. To what extent are aptitudes like intelligence important, or finger dexterity? (11 questions)

4. What temperaments relate to some occupation? (10 questions)

5. What are the physical demands of the job? (6 questions)

6. What physical environment does the job take place in? (7 questions)

7. To what interests does the job relate? (5 questions)

The survey includes clear and detailed instructions on how to answer these questions, making the answers comparable across occupations. In addition, many questions include a grading scale that seemingly targets the possibility of cardinal comparison. Whether cardinal interpretation is appropriate depends on the context, but clearly information is lost when not using these scales in some cardinal fashion. For example, aptitude ratings ask analysts to decide which quintile of the population an occupation falls into. On some questions, however, analysts were only asked to indicate whether they are relevant to a job or not. In each case, we convert the answers provided into numerical values.

Because the questions in the database vary in type and topic, and their number is large, researchers typically make ex-ante decisions on which variables to use. For example, while some questions clearly relate to skills, others clearly do not (such as interest and environment variables). We try to keep any pre-selection to a minimum, and include the three categories of questions on the DOT when we perform Principal Component Analysis. These categories, comprising 18 questions, have in common that they must all be answered on a numeric scale that suggests some form of cardinal interpretation. (This is generally not true for the other categories: they are of the 'yes or no' type.) They also all clearly relate to skills, rather than the environment or personal characteristics of the typical person performing the job. We provide more detail on the 18 questions with our empirical results.

Census

We obtain a crosswalk between DOT occupation codes and Census 1990 occupation codes from the Analyst Resource Center (amongst others associated with the US Department of Labor). Whenever several DOT codes map into one census code, we take the average of component scores as the component score for that Census occupation. This procedure results in 452 occupations.

We use US Census data from IPUMS for all non-skill data (wages, hours worked, employment shares over Census occupations, education, and so forth), where we take a random sample of 50 thousand observations for each of the census years we use. Non-farm hourly wage rates are constructed by combining wage income and non-farm business income, following the example of the Census itself, and correcting for the number of weeks worked and the number of hours worked in a typical week. We reflate all wages to 2012 levels using the 'CPI total items for the United States' from the Federal Reserve Bank of St. Louis. We remove all occupations which are not present throughout our sample, as well as all farm occupations. The final number of occupations for which we have data in our sample is 308.

Population percentiles are obtained as follows. For each occupation in our sample we obtain mean hourly log wages $\overline{w}^{occ}$ and the share of the population employed in the respective occupation x^{occ}. We sort occupations by their mean log hourly wage in 1980. We construct percentile employment shares x^{perc} and average wages $\overline{w}^{perc}$ by mapping the occupation population shares into population percentiles. In particular, we assign to each percentile, the share of each occupation falling into the respective percentile using the 1980 population share per occupation. Doing this, we obtain the following conversion matrix C of dimensions $\#occ \times 100$, which by definition maps the vector of occupation population share vector x^{occ} into population percentile vector x^{perc}:

$$C'_{1980} x^{occ}_{1980} = x^{perc}_{1980} \equiv \mathbf{1}.$$

By construction, the population percentiles obtained in this way are 1 in 1980. Percentile mean wages in 1980 are obtained by multiplying occupational mean wages with the conversion matrix:

$$\overline{w}^{perc}_{1980} = C'_{1980} \overline{w}^{occ}_{1980}.$$

To obtain the change in employment shares between 1980 and 2010, we first calculate how occupational employment changed in terms of 1980 percentiles and then compute the rate of change. In particular, we take the conversion matrix from occupations into percentiles in 1980 and multiply it with the occupation employment share vector in 2010 as follows:

$$\Delta_{2010-1980} x^{perc} = C'_{1980}(x^{occ}_{2010} - x^{occ}_{1980}).$$

This calculation converts 2010 occupational population shares into the 1980 percentile bins. If for an occupation the employment share increased (decreased) relative to 1980, this will result in the respective population percentile to increase (decrease) as well.[6]

[6]Note that by using this strategy, we are restricted to the analysis of occupations which are present in both 1980 and 2010. Thus, we remain silent on the effects of vanishing and newly appearing occupations on the aggregate wage and employment distribution. The procedure follows the approach taken by Autor and Dorn (2013).

For the calculation of changes of wages we similarly multiply the 1980 conversion matrix with 2010 occupational mean log hourly wages and obtain the growth rate by taking the difference of the 2010 and 1980 percentile wages.

1.2.2 Empirics

PCA

The leftmost column of Table 1.1 shows the labels of the 18 DOT questions we include in our analysis. This set is still large, so that we want to reduce it for more tractable empirical analysis. We think of these skills as *ex-ante* equally important indicators of underlying core skills, and want to find out what these underlying skills look like. One method that allows doing so is Principal Component Analysis (PCA).

PCA is a relatively standard technique for dimension reduction, that creates new variables by linear combinations of existing ones. Its objective is to maximize the variance of the new variable, which is called a principle component. Each subsequent component's vector of weights to the variables is assumed orthogonal to the previous ones'. (To make this problem well-defined, variables are first standardized to have mean zero and standard deviation one, and the total weight given to each of them is restricted to be no more than one.) The optimality condition for this problem is a simple eigenvector-eigenvalue decomposition, which yields as many components (eigenvectors) as there are variables, with all components orthogonal to each other. The corresponding eigenvalues relate directly to the variance accounted for by each component. One can simply think of the components as new dimensions: the dimensions are rotated such that the first dimension explains as much variance as possible, thereafter the second, and so on. Thus, the components are identified up to sign and scaling. Variance accounted for by each component are displayed in Figure 1.3. Clearly, the first two components dominate the others in explanatory power: they jointly explain more than 60% of the variance in the data, while no other component explains more than 10%. The third component and further component do not seem to pick up a fundamentally different aspect of skills, but rather seem to modulate the first two. Full PCA results are included in Appendix 1.A.

What do the components look like? Table 1.1 shows the correlation between the first two components with the DOT skill measures over occupations (weighted by their share in employment). Those questions that are negatively correlated with the first component are highlighted. A brief study of the category groups with positive (negative) correlations with the first (second) component unambiguously leads to the conclusion that the first component relates to measures of cognitive ability, while the second component relates to physical skills.

Figure 1.3: PCA Scree Plot

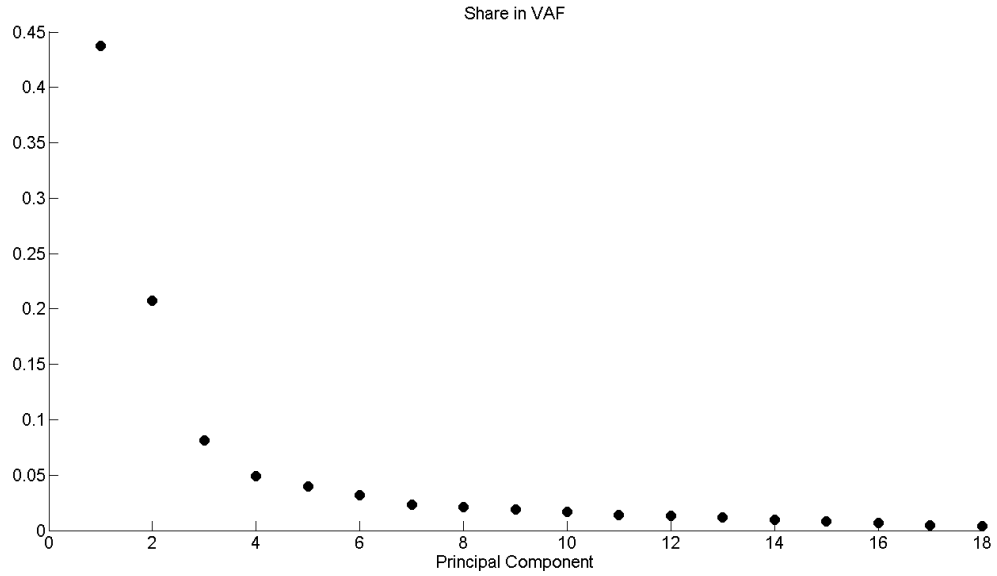


Table 1.1: Component Correlations

Variable	Component 1: 'Cognitive'	Component 2: 'Manual'
Data	0.49	-0.31
People	0.45	-0.27
Things	-0.34	0.12
GED Reasoning	0.47	-0.36
GED Mathematical	0.46	-0.39
GED Language	0.48	-0.38
Specific Vocational Prep.	0.36	-0.22
Intelligence	0.51	-0.44
Verbal	0.51	-0.49
Numerical	0.43	-0.62
Spatial	0.05	0.41
Form Perception	0.05	-0.09
Clerical Perception	0.53	-0.51
Motor Coordination	-0.44	-0.07
Finger Dexterity	-0.63	-0.14
Manual Dexterity	-0.57	0.21
Eye-Hand-Foot Coord.	-0.14	0.44
Color Discrimination	-0.28	0.46

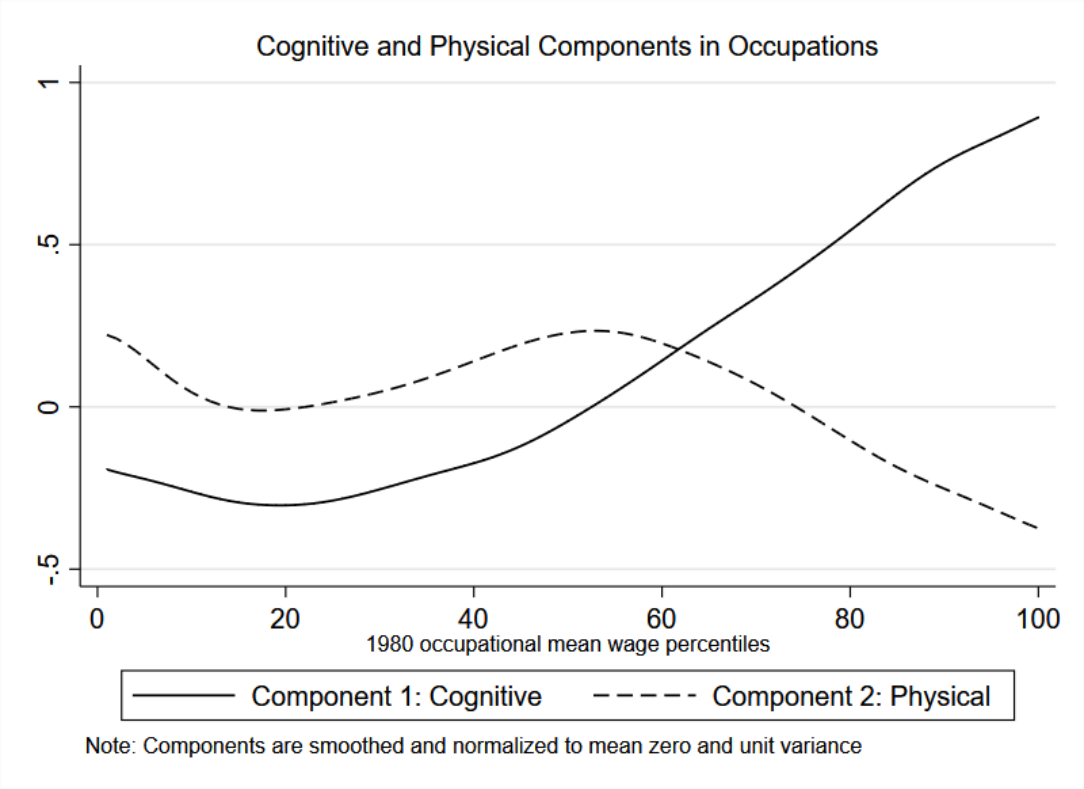
The orthogonality assumption inherent to the method has the natural economic interpretation that these are truly different underlying skills: being good at one does not

mechanically imply being good at another. At the same time, there can certainly still be correlation in abilities in the population of observed occupations. In fact, the correlation between observed occupation scores on the first two components is -0.25: those with more cognitive ability tend to be less able physically, and vice versa.

Covariates

We investigate how the results of the principle component analysis described in the previous relate to the wage distribution. Figure 1.4 plots the first two components over population skill percentiles. As one would expect, the cognitive component is of minor importance in the lower end of the skill distribution, and starts to increase in importance somewhere below the median. In contrast, the physical component is relatively flat for the lower half of the distribution. Above the median it continuously declines with increasing skill level.

Figure 1.4: Components over the Distribution



Taken together, the PCA results imply that the multi-dimensionality of human capital or skills captured in the DOT can be summarized in two main factors, which we call cognitive and manual. As expected, the physical skill is relatively more important in the lower half of the wage distribution, while the cognitive becomes increasingly important for the higher skilled occupations. We view this as an interesting result, as the skilled-versus-unskilled

dichotomy has a long tradition in the analysis of human capital.

Empirically, the skilled-versus-unskilled distinction has often been proxied for by years of education or by comparing college educated workers to those without college education. Figure 1.5 compares the PCA cognitive component to these traditional skill measures. It plots the cognitive component alongside the average years of education and the share of college graduates across wage percentiles in the population. All three measures behave very similarly in a qualitative sense. They are systematically lower for the lower half of the distribution and exhibit a break around the median. In the right half, all three measures are quickly increasing. Table 1.2 presents correlations between these measures of education and the two components, confirming the results from Figure 1.5: Education strongly correlates with cognitive skills, but not with manual skills.

Figure 1.5: Components and Education

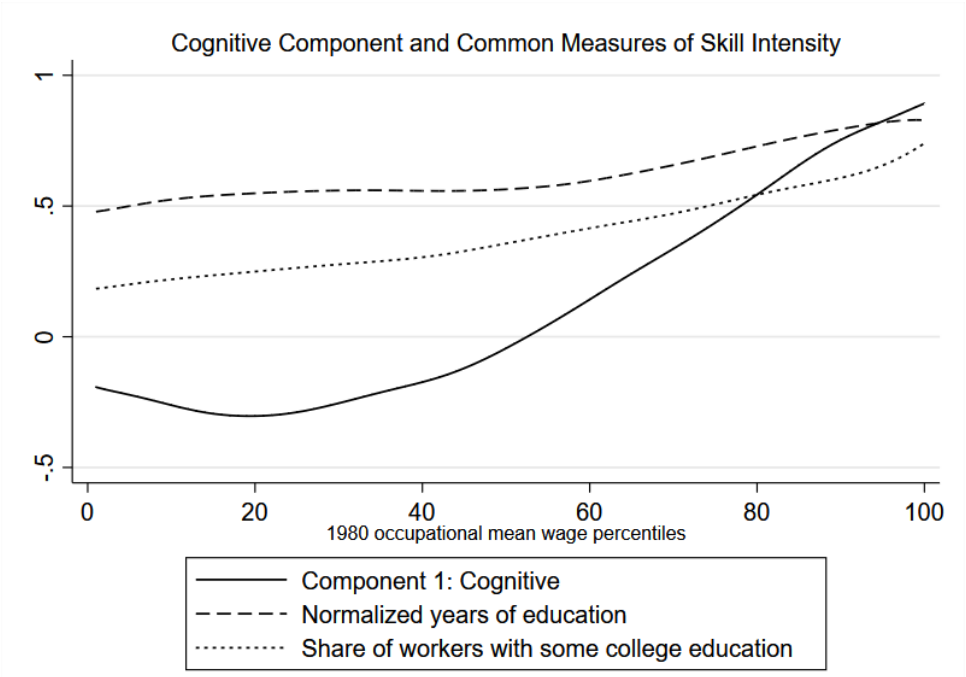


Table 1.2: Correlation between skill components and education measures

Correlations	'Cognitive'	'Manual'
Years of Education	0.97	-0.83
College Share	0.98	-0.70

To sum up, the empirical results suggest that both manual and cognitive skills are important, manual skills predominate in occupations at the bottom of the income distribution while cognitive skills predominate in the upper half. Furthermore, cognitive skills are

highly correlated with traditional measures of schooling, which suggests that the cognitive skill is formed through education.

These observations have implications for the design of our theoretical framework. In particular, they suggest a model of skills where both manual and cognitive skills are important. This is much in line with the literature on Skill-Biased Technological Change, however there is one crucial difference: In the SBTC literature, it is commonly assumed that agents exclusively supply one type of labor, skilled or unskilled. Empirically, the most common strategy for mapping workers to skill types is by applying a cut-off for years of education above (below) which workers are categorized as skilled (unskilled). In contrast, the PCA results presented above suggest that both skills are continuous, i.e. workers in each occupation supply a *bundle* of cognitive and manual skills, rather than just one of the two. This precludes the mapping from years of schooling to skill categories. We will discuss the implications for our theoretical framework in the next Section. Importantly, skill continuity allows for heterogeneity that is not given, but formed endogenously through schooling.

1.3 Model

In the remainder of the paper, we will theoretically and quantitatively explore how tax policy changes distort skill accumulation incentives and thereby impact the shape of the earnings distribution. Motivated by the empirical evidence presented in the previous Section, in this Section we set up a general environment in which earnings are derived from multi-dimensional skills. Individual skills are determined by innate ability and investment to different degrees. We describe the individual income maximization problem trading off time in school and time at work and discuss how progressive taxation distorts this trade-off.

1.3.1 Environment

Human capital accumulation A continuum of agents of mass one derives (pre-tax) earnings from two skills, manual m and cognitive s, quantities of which are measured by $\mathcal{H}_m$ and $\mathcal{H}_s$, respectively. Human capital is accumulated by spending time in school at the beginning of an agent's life. Agents are active for one period of time, $t \in [0, 1]$ and begin their active period in school. They can leave at any time $x \in [0, 1]$ to begin working.[7] Individuals are endowed with cognitive and manual abilities $\alpha = (\alpha_s, \alpha_m)$. Ability α is the only source of heterogeneity in the population. It is continuously distributed with

[7] When we discuss the effects of policy changes in this model, we will essentially be comparing steady states.

pdf $f(\alpha)$ over a finite and positive support $[\underline{\alpha}, \bar{\alpha}]$[2]. Based on the correlation between schooling measures and the cognitive component discussed in Section 1.2, we assume that the manual skill $\mathcal{H}_m$ is innate, i.e. it cannot be accumulated and depends only on manual ability α_m. In contrast, the cognitive skill $\mathcal{H}_s$ is subject to human capital accumulation or schooling. The efficiency of schooling time depends on individual's ability α, the amount of human capital already accumulated and the time spent in school, according to a schooling function $s(\mathcal{H}_s, \alpha, t)$. Cognitive human capital $\mathcal{H}_s(\alpha, x) > 0$, is assumed to be a continuous and twice differentiable function. Finally, we assume that $\frac{\partial \mathcal{H}_s(\alpha_s, x)}{\partial x} > 0$: accumulated human capital is a strictly positive function of time spent in school.

After quitting school, human capital stays constant for the remaining active time of the agents. Human capital accumulation thus follows a differential equation

$$\frac{\partial \mathcal{H}_s}{\partial t} = \begin{cases} s(\mathcal{H}_s, \alpha, t), & t \leq x \\ 0, & t > x \end{cases} \tag{1.1}$$

and the amount of cognitive human capital while working is given by the level of human capital at time x, $\mathcal{H}_s(\alpha, x)$.

Multi-dimensional earnings Individuals are assumed to supply both of their skills to the market simultaneously. Labor supply is assumed inelastic, and set to one. This implies that each individual supplies $\mathcal{H}_m(\alpha)$ units of manual labor and $\mathcal{H}_s(\alpha, x)$ units of cognitive labor when working. Instantaneous gross earnings y of an individual can now be described as

$$y = w_m \mathcal{H}_m(\alpha) + w_s \mathcal{H}_s(\alpha, x). \tag{1.2}$$

Here, w_m and w_s are wage rates for efficiency units of manual and cognitive skills, taken as given by the agents.

On the production side of the economy, final output is produced by an aggregate production function taking the total amounts of manual and cognitive skills in the economy as inputs,

$$Y = F(M, S).$$

Here, M and S are aggregate amounts of manual and cognitive skills in the economy and given by

$$S = \int_{\underline{\alpha}}^{\overline{\alpha}} \mathcal{H}_s(\alpha, x) f(\alpha)\, d\alpha$$

and

$$M = \int_{\underline{\alpha}}^{\overline{\alpha}} \mathcal{H}_m(\alpha) f(\alpha)\, d\alpha = 1,$$

in light of the above normalization of manual skills. We assume competitive input markets, thus wages w_m and w_s are given by their respective marginal products.

1.3.2 Individual Problem

Consumption prices are taken as the *numeraire*. Markets are complete, there are no sources of uncertainty, and a single asset completes the market: agents can save and borrow asset a without limit (except for repayment at $t = 1$) at the discount rate: $r = \tilde{r}$.[8] Taxes are assumed to be paid instantaneously over the rate of income y. This is a crucial model ingredient: in reality, tax schedules are applied yearly, which is a short frequency compared to the length of the life cycle. In our continuous time model, we capture this by applying the tax schedule to the wage rate at any instance.

An individual's problem then looks as follows:

$$\max_{\substack{x\in[0,1],\\ \{c_t\}_{t\in[0,1]}}} \quad \int_0^1 e^{-\tilde{r}t}\frac{c^{1-\sigma}}{1-\sigma}dt \tag{1.3}$$

$$\begin{aligned}
&\text{subject to} \quad \forall t: \\
&\frac{\partial a_t}{\partial t} = -c_t(1+\tau_c) + a_t r \quad \text{if} \quad t \leq x, \\
&\frac{\partial a_t}{\partial t} = y_t(1-\tau_y(y_t)) - c_t(1+\tau_c) + a_t r \quad \text{if} \quad t > x, \\
&a_0 = 0, \quad a_1 \geq 0, \\
&c_t \geq 0.
\end{aligned}$$

Agents decide on the duration of their education and on the life-cycle profile of consumption and savings.

The government levies taxes on consumption and earnings, τ_c and $\tau_y(\cdot)$, in order to meet wasteful government spending target G. We assume that the earnings tax $\tau_y(\cdot)$ is governed by two parameters, responsible for average tax level ϕ and the degree of tax progressivity θ, $\tau_y(\cdot) = \tau_y(\cdot\,;\phi,\theta)$. An specific example of such an earnings tax function is $\tau_y(y) = 1 - \frac{1}{\phi}(y)^{-\theta}$. This is the tax function we will use in our quantitative exercise in Section 1.6.

1.4 Tax Policy, Inequality Changes and Polarization

In this Section, we study the effect of changes in the tax policy, in particular changes to the tax progressivity θ on the shape of the earnings distribution. Investing in education enables individuals to achieve higher earnings in a shorter time span. Recall that in our framework, taxes are not applied to life-time income, but instead levied on instantaneous

[8] We do not model the capital stock of the economy in general equilibrium, which yields the same results as an economy that is 'small' and open to capital only, or with an aggregate production technology that is linear in capital.

earnings, in order to resemble real-world income taxation. For a given life-time income, positive tax progressivity punishes higher per-period earnings relative to an earnings profile that spreads out lower earnings over a larger fraction of the life-cycle. Through this channel, tax progressivity directly influences the optimal choice of time in school, x^* and cognitive human capital, $\mathcal{H}_s(\alpha, x^*)$. Next, we will present a static version of the general framework above. We will then define the notions of inequality and polarization in this framework and provide conditions for the latter to arise in response to a policy change.

1.4.1 A simple static framework

Life-cycle problem (1.3) above has a straightforward solution. Since individuals can only decide between going to school or working full-time, the amount of human capital is fully determined by the time spent in school, x, and cognitive ability, α_s. Second, because markets are complete, agents smooth consumption and the choice of the optimal x is unconstrained. As there are no other choices in the model, the agent now simply maximizes lifetime after-tax income with respect to time in school:

$$\begin{aligned} \max_{x \geq 0} \quad & (1-x)y(1-\tau(y;\phi,\theta)) \\ & \text{subject to:} \\ & y = w_m + w_s \mathcal{H}_s(\alpha, x). \end{aligned} \tag{1.4}$$

For tractability, we assume that there is no heterogeneity in the manual skill, so that we can normalize its level to one, $\mathcal{H}_m = 1$. This implies that the aggregate amount of manual skill, M is also equal to one.

In the following, we will be interested in how changes in the tax progressivity θ shape the income distribution in this environment. As discussed, the optimal schooling decision will be directly governed by the degree of tax progressivity. From now on, we will therefore directly work with $h(\alpha, \theta) \equiv \mathcal{H}_s(\alpha, x^*)$ instead of human capital $\mathcal{H}_s(\alpha, x)$, where x^* is the argmax of individual income maximization problem (1.2).

1.4.2 Policy changes and earnings distribution

Our basic interest is the study of the effect of taxation on the earnings distribution in the presence of multi-dimensional skills. To this end, we will formalize the notion of earnings inequality and earnings polarization in our framework and provide conditions for either of them to arise. As will become clear below, earnings polarization is a special case of earnings inequality growth, with additional restrictions on the relative movements of earnings within the lower tail of the distribution.

As outlined above, we will link the changes in the earnings distribution reported by

Autor and Dorn (2013) to changes in the tax incentives to relative skill supply studied by Guvenen et al. (2014). Thus, we will be interested in the relative (percentage change) effects of policy on earnings, since this is the theoretical equivalent of Figure 1.2. In particular, we will then be interested in how these policy effects differ across different parts of the income distribution. Our general object of interest is therefore given by

$$\frac{\frac{\partial y}{\partial \theta}}{y}.$$

It turns out, this can be easily decomposed into separate parts as follows:

$$\begin{aligned}
\frac{\left(\frac{\partial y}{\partial \theta}\right)}{y} &= \frac{\left(\frac{\partial\left(\frac{y}{w_s}\right)}{\partial \theta}\right)}{\left(\frac{y}{w_s}\right)} + \frac{\left(\frac{\partial w_s}{\partial \theta}\right)}{w_s} \\
&= \frac{\left(\frac{\partial\left(\frac{w_m}{w_s}\right)}{\partial \theta}\right) + \frac{\partial h}{\partial \theta}}{\left(\frac{w_m}{w_s}\right) + h} + \frac{\left(\frac{\partial w_s}{\partial \theta}\right)}{w_s} \\
&\equiv \underbrace{\frac{w'_\theta}{(w+h)}}_{\text{price effect}} + \underbrace{\frac{h'_\theta}{(w+h)}}_{\text{quantity effect}} + \underbrace{\frac{w'_{s\theta}}{w_s}}_{\text{level effect}}
\end{aligned} \tag{1.5}$$

In the above, percentage changes in income have been separated into three terms. The first two terms describe the potential trade-off policy created in a multi-dimensional model: on the one hand, policies can increase (or decrease) incentives to acquire human capital, which we call a *quantity effect*, but when they do so for all individuals this increases (decreases) the overall supply of learnable skills in the economy, which can decrease (increase) their relative price - a *price effect*. Both these terms would then move in the same direction, but their relative importance and strength depends on an individual's schooling responsiveness to policy. This responsiveness will in principle depend on the level of the ability parameter α, generating potentially non-linear effects of policy changes on income changes. The last term above affects all individuals equally in percentage terms. It arises because wage effects are described in skill premium terms, but a policy reform can also impact the overall productivity level in an economy - hence the name *level effect*.

Polarization in this environment arises if relative income changes in response to a policy change are stronger in the tails of the income distribution than in the center of the distribution. Since in the model, income is entirely determined by ability, this is equivalent to comparing income responses for different ability levels. Formally, inequality growth and polarization in response to tax policy changes can be defined in terms of relative income changes as follows.

Definition 1. *Inequality Change and Polarization. Inequality change exists in response*

to a policy change in θ if for $\underline{\alpha}$ and $\overline{\alpha}$ the following inequality holds:

$$\left.\frac{\left(\frac{\partial y}{\partial \theta}\right)}{y}\right|_{\alpha=\overline{\alpha}} < \left.\frac{\left(\frac{\partial y}{\partial \theta}\right)}{y}\right|_{\alpha=\underline{\alpha}} < 0. \tag{1.6}$$

Polarization exists if in addition to equation (??), for some $\hat{\alpha} \in (\underline{\alpha}, \overline{\alpha})$ the following holds as well:

$$\left.\frac{\left(\frac{\partial y}{\partial \theta}\right)}{y}\right|_{\alpha=\underline{\alpha}} < \left.\frac{\left(\frac{\partial y}{\partial \theta}\right)}{y}\right|_{\alpha=\hat{\alpha}} < 0. \tag{1.7}$$

Definition 1 restates income inequality growth and polarization in response to a decline in tax progressivity θ in concise terms. First, the inequality aspect, i.e. high income individuals pulling away even further from the rest of the population, requires a stronger relative income response to a policy change for high ability individuals than low ability individuals. Since this effect is negative for higher levels of tax progressivity θ, the response will be more negative for high ability individuals. Second, the non-monotonicity in the lower tail of the income distribution distinguishes polarization from general trends in overall inequality: low-income individuals are able to partially catch-up to medium income individuals, while overall inequality still increases.

1.4.3 Conditions for Inequality Growth and Polarization

Our reformulation of the life-cycle problem (1.3) as income maximization problem (1.2) allows us to establish conditions for polarization to arise in our framework. After presenting those conditions, we will show their sufficiency for polarization to arise. The intuition for this result can also be seen from reformulating the decomposition (1.5) in terms of tax policy elasticities as follows:

$$\varepsilon_\theta^y = \varepsilon_\theta^w \frac{w}{w + h(\alpha)} + \varepsilon_\theta^h(\alpha) \frac{h(\alpha)}{w + h(\alpha)} + \varepsilon_\theta^{LE}. \tag{1.8}$$

Equation (1.8) rewrites the total relative earnings elasticity in terms of weighted elasticities of the price and the quantity effect, both of which are functions of α and the level effect ,which is independent of α. Polarization can arise because the elasticity of the price effect has the same size for all abilities, while the elasticity of the quantity effect potentially grows in α. In addition, the weights also change in ability, since for higher ability the share of income generated from cognitive human capital increases. Depending on the shape of the change of elasticity of the quantity effect, we can have non-monotone changes in relative income across abilities. In particular, if $\varepsilon_\theta^h(\alpha)$ is small in absolute magnitude for small and medium α, the growing weight on the second, quantity elasticity term

may initially decrease the absolute magnitude of overall elasticity as we move along the earnings distribution. Only once $\varepsilon_\theta^h(\alpha)$ is large enough in absolute magnitude will the absolute magnitude of the overall elasticity begin to grow in earnings.

In the following, we will make this reasoning more precise by first laying out the assumptions sufficient for polarization to arise, and then go through the precise mechanism. The main purpose of this exercise is to clarify and formalize the intuition just laid out.

Definition 2. *Define $\hat{\alpha} \in (\underline{\alpha}, \overline{\alpha})$ as any α that generates a local extremum in the tax elasticity of income,*

$$\left.\frac{\partial \varepsilon_\theta^y(\alpha)}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = 0.$$

Inequality change and polarization in Definition 1 were defined for an arbitrary interior $\hat{\alpha}$. Definition 2 restricts, as we will see below, the interior $\hat{\alpha}$ to the ability level that under the below assumptions minimizes (in absolute terms) the income elasticity $\varepsilon_\theta^y(\alpha)$.

Assumption 1. *Shape of the human capital elasticity. Human capital elasticity $\varepsilon_\theta^h(\alpha)$ behaves relative to the relative wage elasticity ε_θ^w as follows for different ability levels:*

- *The human capital quantity elasticity is increasing and convex in ability level α in absolute terms, $\frac{\partial \varepsilon_\theta^h(\alpha)}{\partial \alpha} < 0$ and $\frac{\partial^2 \varepsilon_\theta^h(\alpha)}{\partial \alpha^2} < 0$, and at the lower bound approximately zero: $\frac{\partial \varepsilon_\theta^h(\underline{\alpha})}{\partial \alpha} \approx 0$.*

- *For abilities $\alpha \leq \hat{\alpha}$, the human capital elasticity is lower in absolute terms than the elasticity of relative prices, $\varepsilon_\theta^h(\alpha) - \varepsilon_\theta^w > 0$.*

- *For high ability individuals, the human capital elasticity is higher in absolute terms than the elasticity of relative prices, $\varepsilon_\theta^h(\overline{\alpha}) - \varepsilon_\theta^w < 0$.*

Assumption 1 states that for low ability individuals, the relative price elasticity is stronger than the quantity elasticity. The quantity elasticity is increasingly growing in ability and for abilities high enough, it becomes larger than the relative price elasticity in absolute terms.

Assumption 2. *Human capital function. Human capital accumulation is strictly convex and positive in α, $h'_\alpha, h''_\alpha > 0$. In addition, for all $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, the following restriction on the shape of the optimal human capital quantities holds:*

$$\frac{1}{w+h} < \frac{2(h'_\alpha)^2}{h''_{\alpha\alpha}}.$$

Recall from above that $h(\alpha, \theta)$ is the cognitive human capital resulting from the optimal schooling decision of the agent, $h(\alpha, \theta) \equiv \mathcal{H}_s(\alpha, x^*)$. Therefore, Assumption 2 effectively imposes restrictions shape of the schooling technology. In particular, it is required that

the human capital is convex in ability, but cannot increase too quickly – h''_α has to be sufficiently small.

Result 1. *Given Assumption 1, inequality changes as defined in Definition 1 occur in response to a change in tax policy θ. If in addition Assumption 2 also holds, polarization as defined in Definition 1 occurs as well.*

Proof: For the inequality change part, we will show that income elasticity $\varepsilon^y_\theta(\alpha)$ is larger in absolute terms for $\overline{\alpha}$ than for $\underline{\alpha}$. For the polarization part, we will show that under Assumption 2, there is a unique $\hat{\alpha}$ as defined in Definition 2 and this $\hat{\alpha}$ is the argmax of the maximum of $\varepsilon^y_\theta(\alpha)$ (minimum in absolute terms).

Inequality change: To show that $0 > \varepsilon^y_\theta(\underline{\alpha}) > \varepsilon^y_\theta(\overline{\alpha})$, we will first show that $\varepsilon^y_\theta(\underline{\alpha}) - \varepsilon^{LE}_\theta > \varepsilon^w_\theta$ and second that $\varepsilon^y_\theta(\overline{\alpha}) - \varepsilon^{LE}_\theta < \varepsilon^w_\theta$. To show the former, consider

$$\begin{aligned}
&\varepsilon^y_\theta(\underline{\alpha}) - \varepsilon^{LE}_\theta > \varepsilon^w_\theta \\
\Leftrightarrow\, &\varepsilon^w_\theta \frac{w}{w+h(\alpha)} + \varepsilon^h_\theta(\alpha)\frac{h(\alpha)}{w+h(\alpha)} > \varepsilon^w_\theta \\
\Leftrightarrow\, &\varepsilon^w_\theta \frac{w}{w+h(\alpha)} + \varepsilon^h_\theta(\alpha)\frac{h(\alpha)}{w+h(\alpha)} > \varepsilon^w_\theta \frac{w}{w+h(\alpha)} + \varepsilon^w_\theta \frac{h(\alpha)}{w+h(\alpha)} \\
\Leftrightarrow\, &\varepsilon^h_\theta(\underline{\alpha}) > \varepsilon^w_\theta,
\end{aligned}$$

where the last inequality holds by Assumption 1. For the high ability case $\varepsilon^y_\theta(\overline{\alpha}) - \varepsilon^{LE}_\theta < \varepsilon^w_\theta$, a similar argument holds. Together, this implies that $\varepsilon^y_\theta(\underline{\alpha}) - \varepsilon^{LE}_\theta > \varepsilon^w_\theta > \varepsilon^y_\theta(\overline{\alpha}) - \varepsilon^{LE}_\theta$. $\varepsilon^y_\theta(\underline{\alpha}) > \varepsilon^y_\theta(\overline{\alpha})$ is implied by the last inequalities, establishing inequality change as defined in equation (1.6) from Definition 1.

Polarization: To show that the unique global maximum of $\varepsilon^y_\theta(\alpha)$ is at $\hat{\alpha}$, we show first that $\hat{\alpha}$ is the only extremum, and second that the first derivative is strictly larger (smaller) than zero for all α smaller (larger) than $\hat{\alpha}$. The first derivative of $\varepsilon^y_\theta(\alpha)$ with respect to α is given by

$$\frac{\partial \varepsilon^y_\theta(\alpha)}{\partial \alpha} = \left(\varepsilon^h_\theta(\alpha) - \varepsilon^w_\theta\right)\frac{w h'_\alpha}{(w+h)^2} + \frac{\partial \varepsilon^h_\theta(\alpha)}{\partial \alpha}\frac{h}{w+h}. \tag{1.9}$$

Therefore, $\frac{\partial \varepsilon^y_\theta(\alpha)}{\partial \alpha} > (<)\, 0$ boils down to

$$\left(\varepsilon^h_\theta(\alpha) - \varepsilon^w_\theta\right)\frac{w h'_\alpha}{(w+h)^2} > (<)\, -\frac{\partial \varepsilon^h_\theta(\alpha)}{\partial \alpha}\frac{h}{w+h}.$$

For the left-hand side, Assumption 1 implies that $\left(\varepsilon^h_\theta(\alpha) - \varepsilon^w_\theta\right)$ is strictly declining in α, positive for $\underline{\alpha}$ and negative for $\overline{\alpha}$. Define $\tilde{\alpha}$ as the α such that $\varepsilon^h_\theta(\tilde{\alpha}) - \varepsilon^w_\theta = 0$. Note that by Assumption 1 $\overline{\alpha} > \tilde{\alpha} > \hat{\alpha}$ holds. Assumption 2 implies that $\frac{w h'_\alpha}{(w+h)^2}$ is strictly decreasing in α and strictly positive. Together this implies that the left-hand side is strictly declining in α for $\alpha < \tilde{\alpha}$, positive for $\alpha < \tilde{\alpha}$ and negative for $\alpha > \tilde{\alpha}$. For the right-hand side,

Assumption 1 implies that $-\frac{\partial \varepsilon_\theta^h(\alpha)}{\partial \alpha}$ is ≈ 0 for $\underline{\alpha}$ and strictly increasing and positive for all $\alpha > \underline{\alpha}$.

Taken together, this implies that $\frac{\partial \varepsilon_\theta^y(\underline{\alpha})}{\partial \alpha} > 0$ and $\frac{\partial \varepsilon_\theta^y(\overline{\alpha})}{\partial \alpha} < 0$. Furthermore, since the left-hand side is strictly declining while positive and the right-hand side strictly increasing and positive, there exists exactly one α for which $\frac{\partial \varepsilon_\theta^y(\alpha)}{\partial \alpha} = 0$. This proves the existence of a unique $\hat{\alpha}$ as defined in Definition 2. Since for all $\alpha < \hat{\alpha}$, we have that $\frac{\partial \varepsilon_\theta^y(\alpha)}{\partial \alpha} > 0$ and for all $\alpha > \hat{\alpha}$, we have that $\frac{\partial \varepsilon_\theta^y(\alpha)}{\partial \alpha} < 0$, $\hat{\alpha}$ is a global maximum. Since $\hat{\alpha} \in (\underline{\alpha}, \overline{\alpha})$, this establishes polarization as defined in equation (1.7). $\square$

The aim of this Section has been to detail conditions on the price and quantity effects for earnings polarization to arise in our framework with two types of skills and general equilibrium price effects. We show that depending on the shape of the elasticities, polarization can arise in our framework of two-dimensional skills and general equilibrium skill price effects. In the following, we will first present some reduced-form cross-country evidence for our mechanism in the next Section, and then try to quantify the economic importance of this supply-side channel in a richer version of our model in Section 1.6.

1.5 Models versus Data

1.5.1 Across Countries

Tax systems differ in progressivity across countries (Figure 1.6). Our model makes clear predictions on the role of progressivity in income inequality: more progressive tax systems produce less inequality as measured by relative earnings in the income distribution. This is driven by changes in the upper half of the income distribution, while inequality in the bottom half varies little with tax progressivity. Neither prediction is made by a model with one-dimensional skills, nested in our setup.[9] We now investigate these predictions in cross-country data, for which we need measures of tax progressivity and relative earnings inequality.

Coen-Pirani (2017) sets forth a method to obtain measures of tax progressivity from OECD data, which works as follows: If we assume that both gross and net earnings are log-normally distributed and that taxes follow the functional form assumed above, measured Gini coefficients of gross and net earnings can be used to back out an estimate of θ. A panel data set of Gini coefficients is available from the OECD Income Distribution Database. We use data on the working age population (ages 18 to 65), using the income

[9] We think of tax systems as exogenous to the remainder of the economy. If tax progressivity is in some way a response to higher earnings inequality, this would counteract our mechanism and make it harder to find a correspondence between theory and data.

definition that the OECD followed until 2011 for better availability and comparability of data. Data are available for about 30 OECD member countries, covering a period from the mid 1970s to 2015. Coverage is thin for earlier years, but improves towards the end of the sample. Because the panel is rather unbalanced, we average the resulting measures of tax progressivity for the years 2010–2015, and use this as a cross-section of country-level tax progressivity.

Also available from the OECD is an unbalanced panel of relative earnings inequality measures across countries and over time. The underlying population are full-time employees of either gender. These include the earnings ratio of the 90th percentile cut-off to the 50th percentile cut-off, or *90-50 ratio*, and the same for the 50th and 10th percentile, the *50-10 ratio*. While these two measures describe relative inequality above and below the median, the resulting *90-10 ratio* measures inequality. We choose to use these measures because their movement has a close correspondence to what we consider in our theoretical exposition: if relative (percentage) changes are the same across the distribution, then these measures will remain unchanged with tax progressivity. We again average over the years 2010–2015. The overlap between the two datasets consists of 32 countries.

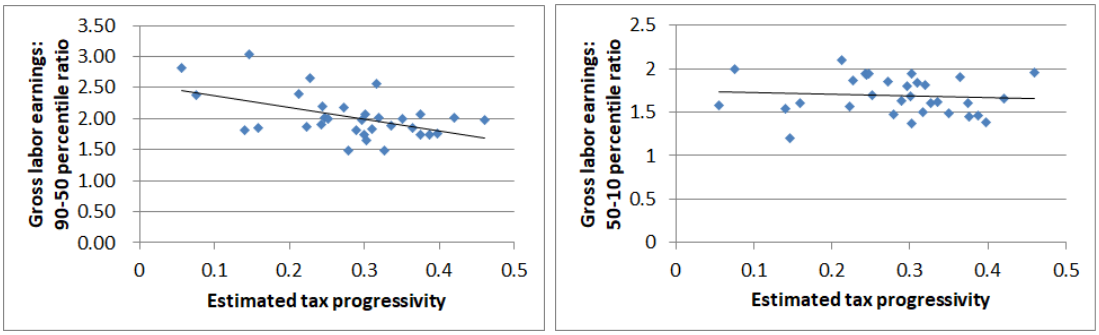
Next, we estimate the (linear) impact of tax progressivity on earnings inequality at different points in the distribution. Results of OLS regressions of the latter on the former are displayed in Table 1.3. Figure 1.6 presents the results graphically. Tax progressivity is generally associated with a reduced relative earnings inequality. For the 90-10 ratio, the slope is statistically significant at the 5% level. For the 90-50 ratio the slope is statistically significant at the 1% level. For the 50-10 ratio, the slope is not statistically significant, even at the 10% level. While tax progressivity has quite some explanatory power in the upper half of the distribution, as measured by the R^2, this is not true for the lower half of the distribution.

Table 1.3: Regression results

	Inequality measure		
	90-10 ratio	90-50 ratio	50-10 ratio
Progressivity ϕ_1	-1.17	-1.92	-0.19
	(0.54)	(0.59)	(0.41)
Constant	Yes	Yes	Yes
Observations	32	32	32
R^2	0.14	0.26	0.01

All these results align very well with our model prediction, even as we are looking at simple linear relations. In a more extensive quantitative exercise, Guvenen, Kuruscu, and Ozkan (2014) analyze the responses of a one-dimensional model of human capital to changes in

Figure 1.6: Tax progressivity and Inequality across Countries



tax progressivity. Their model has several added features, such as flexible labor supply, and a more flexible functional form for average tax rates. While they show that their model does well in accounting for 90-10 ratios, it is less succesful in disentangling 50-10 ratios. Our analysis suggests that this is due to the multi-dimensional nature of skills, which is most relevant to the bottom half of the distribution. It also suggests that the productivity effect of taxes on human capital may be mitigated by general equilibrium effects.

1.5.2 Over Time

Tax progressivity in the United States has declined dramatically since the 1970s (Figure 1.1). That same observation applies to many other countries (Guvenen, Kuruscu, and Ozkan, 2014). What implications would this have had for other observables, in particular changes to the shape of the earnings distribution? The literature review above already shows that the earnings distribution is shaped by a number of different forces. That fact significantly limits the extent to which we can verify the direct impact of tax changes on inequality through our mechanism empirically. Nevertheless, we attempt to provide a qualitative discussion.

Educational decisions are decisions for the long run. Agents expectations of future policies are therefore key to the empirical mechanism we describe, and observed transitions may be slow. In any case, one would expect that younger cohorts react more strongly to incentives than older ones, so that empirically it should be the younger cohorts that cause polarization. This is indeed what the empirical literature finds. Cortes, Jaimovich, and Siu (2017) document that the fall in what they call 'routine' occupations in the middle of the distribution can be largely attributed to two groups: young and prime-aged men with low levels of education where it comes to 'routine manual' occupations, and young and prime-aged women with intermediate levels of education where it comes to 'routine cognitive' occupations. In terms of age structure, this lines up well with the implications of our mechanism. While our model does not speak to gender *per se*, the gender differences

these authors highlight underline our main suggestion for further research: changes in labor market discrimination may be important. We will come back to this in more detail in the concluding Section 1.7.

Implications of changes in tax incentives for wages are summarized by rising inequality, in gross wages and even more so net of taxes, and polarization. While these phenomena can also be observed in the data, their underlying components cannot. This is because observed wages are the product of human capital quantities and prices. Did wage inequality grow due to greater differences in human capital or due to rising prices for the highly skilled? Our mechanism would suggest the former. The theory of Skill-Biased Technological Change on the other hand takes growth of educational attainment at face value as a measure of human capital quantities, and interprets its slow-down as a reason for rising prices for the highly skilled.

Separating human capital quantities from their price is a central empirical challenge in the labor literature and existing evidence is scarce. One approach is to identify an age in the life-cycle at which human capital is unlikely to change much, and attribute wage changes at that age to changes in the price of human capital. This is the approach followed by Bowlus and Robinson (2012). These authors do not find large changes in prices at all, attributing changes in the wages of different educational groups to changes in human capital. This would be more in line with our mechanism than for example SBTC, although to cause polarization on its own our general equilibrium effect would require a growing relative price of manual versus cognitive skills. Price estimates by such skill types are unfortunately unavailable.

Similar caveats apply to direct measures of human capital (such as schooling attainment), measured skill premia, and before and after tax returns to schooling. While our model makes predictions for each of these, it is not clear what is the relevant empirical counterpart. A number of possible comparisons are further complicated by the fact that our model is not a growth model, so that it cannot account for longer-run trends in these data.

Finally we return to our initial comment: our mechanism is unlikely to have been the only relevant change during the period. Other explanations focus on secular technological developments that have shaped the wage distribution through changes in labor demand. These explanations are complementary to ours as long as relative prices of skills move in the same direction as in our model. That holds for the literature that describes how middling 'routine' occupations are more prone to automation. The same applies to papers that explain the growth of service occupations at the bottom of the distribution through changes in demand. SBTC fits our model less well, since it starts with the assumption that it is prices of human capital that have caused inequality to grow. Future research

will hopefully shed further light on this debate.

After having discussed evidence for some general predictions of our framework, next we will present a quantitative version and subsequently use it for a more formal investigation of the quantitative relevance of our mechanism.

1.6 An Enriched Model

In this Section, we extend our model to include heterogeneity in manual (non-learnable) skills and choose some functional forms. We then parameterize our model to reproduce several key stylized facts of the US economy, and use it to evaluate counter-factual policies.

1.6.1 Model Description

A continuum of agents, whose total mass equals one, live for $t \in [0, 1]$, first goes to school until $t = x$ and then works. When in school ($x \leq t$), individuals build learnable human capital according to the following law of motion:

$$\frac{\partial h_{s,t}}{\partial t} = \beta t^{\beta-1} \alpha_s h_{s,t}. \tag{1.10}$$

Thereafter, $\frac{\partial h_{s,t}}{\partial t} = 0$. This function resembles more conventional human capital functions such as the one due to Ben-Porath (1967), but the time-in-school structure keeps the model computationally simple. Time in school is more productive for the more able and educated, but diminishes over time. $h_{s,0}$ is assumed linear in α_s, so that the two are perfectly correlated. This simplifies the problem significantly at little cost. Non-learnable human capital is given by $h_{m,t} = h_{m,0} = \alpha_m$. Both skills are assumed to be independently drawn from normal distributions (winsorized at three standard deviations from the mean), resulting in a tuple (α_m, α_s) for each individual. When working ($x > t$), individuals derive income from both types of human capital:

$$y_t = w_m h_{m,t} + w_s h_{s,t}. \tag{1.11}$$

From here on out, the individual problem is the same as in equation (1.3) above. We consider overlapping generations such that the population distribution is always in steady state. Let the distribution of type tuples $(\alpha_m, \alpha_s) \in \mathcal{A}$ be denoted by λ. Define human capital aggregates as follows (where $I_{[\cdot]}$ is an indicator function):

$$H_m = \int_0^1 \int_{\mathcal{A}} h_{m,t} I_{[t>x]} \, d\lambda \, dt \tag{1.12}$$

$$H_s = \int_0^1 \int_{\mathcal{A}} h_{s,t} I_{[t>x]} \, d\lambda \, dt. \tag{1.13}$$

Aggregate production takes place using the following production function:

$$Y = F(H_m, H_s) = A\Big[\gamma H_m^{\rho} + (1-\gamma)H_s^{\rho}\Big]^{\frac{1}{\rho}}. \tag{1.14}$$

The elasticity of substitution between the two inputs is given by $\frac{1}{1-\rho}$, and γ is a share parameter. We normalize output so that $A = 1$.

A government sets taxes τ_c and $\tau_n(\cdot)$. Its budget is balanced by expenditures G that are assumed not to influence any of the above:

$$\int_0^1 \int_{\mathcal{A}} c_t \tau_c + y_t \tau_n(y_t) I_{[t>x]} \, d\lambda \, dt = G. \tag{1.15}$$

Definition 3. *An equilibrium of the model is defined as:*
Wages w_m, w_s,
allocations H_m, H_s,
government spending G,
decision rules for x, $\{c_t\}_{t\in[0,1]}$ $\quad \forall \quad (\alpha_m, \alpha_s) \in \mathcal{A}$
such that given the parameters of the model the following holds:

- *individual decision rules solve problem (1.3)*
- *goods markets clear:*

$$Y = \int_0^1 \int_{\mathcal{A}} c_t \, d\lambda \, dt \tag{1.16}$$

- *labor markets clear (equations (1.12) and (1.13))*
- *wages equal marginal products (of equation (1.14))*
- *and the government budget constraint is balanced (equation (1.15)).*

1.6.2 Parameterization

Equilibria of the economy are found numerically. Parameters are set to match moments of the data in the early 2000s. In doing so, the following parameterizations of initial abilities and human capital stocks is used. Let $\tilde{\alpha}_s$ denote a standard normal distribution, winsorized at three standard deviations.

$$\alpha_s = \mu_s + \sigma_s \tilde{\alpha}_s, \tag{1.17}$$

$$h_{s,0} = 1 + (\tilde{\alpha}_s - \underline{\tilde{\alpha}}_s)\psi_s. \tag{1.18}$$

$\underline{\alpha}_s$ is the lowest level of α_s. The lowest level of $h_{s,0}$ is normalized to 1, while average learning ability, the spread in learning ability, and the spread in initial learnable human capital is controlled by parameters. Likewise,

$$h_{m,0} = 1 + (\alpha_m - \underline{\alpha}_m)\psi_m, \tag{1.19}$$

where α_m is standard normal and ψ_m controls the spread of initial non-learnable human capital.

Table 1.4: Parameters and moments

Parameter	Value	Moment	Model	Data
σ	2.857	Elasticity of intertemporal substitution	0.350	0.350
ψ_m	0.141	Earnings variance at start of working life versus overall	0.528	0.500
ψ_s	28.068	Gini coefficient of gross earnings	0.346	0.440
μ_s	0.947	Average share of working age spent in school	0.030	0.034
σ_s	0.225	Variance of share in school	0.002	0.002
β	0.858	Share with zero education after age 18	0.478	0.456
ρ	0.286	Elasticity of substitution in production	1.400	1.400
γ	0.519	Non-learnable share of output	0.248	0.250

Table 1.4 reports data moments. Some of our model parameters are straightforwardly informed by moments of the data, while for others much less clear-cut measures are available. We use the midpoint of the range of elasticities of intertemporal substitution reported in Havránek (2015) to set the same in the model (σ), but that parameter does not influence any of the results we report. The spread of both initial human capitals is important for overall earnings variation, and their relative size helps determine the extent to which that variation is present at age 0. Thus, we target the Gini coefficient of gross earnings as reported by the OECD for the year 2000. We also target a ratio of earnings variance at age 0 versus earnings variance overall of 1/2. While we do not have a precise estimate for this number from the data, research using the life-cycle of earnings Huggett, Ventura, and Yaron (2011) suggests about two thirds of earnings are pinned down *after* tertiary education. Finally, to determine the average and spread of ability, we target the share of a potential 48 years of working life from age 18 that is spent in school (i.e. college and beyond), the variance of these shares, and the share of pupils who do not spend any time in college. We calculate the data moments from the 2000 Census sample described in the above, where all education beyond 12th grade is counted as taking place during the adult life cycle.

Finally, the parameters in the production function are key to size general equilibrium responses. Unfortunately, no reduced form results on general equilibrium effects between skills as we describe them are available. Instead, we rely on evidence on general equilibrium effects between college educated and non-college educated labor. Here, a large body of evidence suggests an elasticity of substitution of about 1.4 (see for example Katz and Murphy (1992) and Ciccone and Peri (2005)). Because these two groups would both use either type of human capital, we take the view that this is a very conservative estimate of the two elasticity of substitution that is relevant to our model. To tie down the share

parameter of the production technology, we target the share of non-learnable human capital in output. Again, no direct evidence is available, so that we tentatively set this target to 25%.

Consumption taxes are set to 7.5%, following the 2003 Figure reported in McDaniel (2007). We estimate the tax function used in the above from tax rates at different levels of average US earnings for 2003, and then do the same for 1983, following Guvenen, Kuruscu, and Ozkan (2014) (we use the same data as those authors). This results in an estimate $\phi_1 = 0.119$ for 2003, which is used for parameterizing the model, and an estimate of $\phi_1 = 0.188$ for 1983, which we use in our counter-factual analysis below. ϕ_0 is set to clear the government's budget constraint.

Table 1.4 also demonstrates the model's ability to match the data. Overall, model moments are close to data moments, although the model does struggle to create sufficient earnings heterogeneity to match the economy's inequality levels.

1.6.3 Results

To analyze the results of tax progressivity, we compare the steady state earnings distribution of the 1983 estimate of ϕ_1 to the steady state distribution with the 2003 estimate. We think of this as a counter-factual reform in which tax progressivity was reduced. The procedure yields a reform that is per definition realistic, both in shape and magnitude. We would not want to argue that our results are empirical in the sense that they have bearing on the change in the period. (For that to be the case, one would want to consider other factors, as well as the transition from one steady state to another.) Rather, we are looking for a counter-factual experiment that gives us a feeling for the effect sizes in our model.

We then turn to measures of inequality. Indeed, reducing the progressivity parameter has increased the 90-10 ratio about one-for-one, which is what we also find in our cross-country analysis. This increase can be almost entirely attributed to the upper half of the distribution i.e. the 90-50 ratio. Again, this is entirely in line with our cross-country findings. These results give us confidence that the model adequately captures the reaction of the earnings distribution to tax progressivity.

Figure 1.7a shows the results graphically (labeled 'baseline'). It is apparent that some polarization occurs, but little: the bottom wages grow a few tenths of percent more than those with the lowest wage growth. The top grows by almost 7% more than the lowest point.[10] To bear out polarization given the large increase in inequality in the top half, we

[10]For those interested, we report that this is 6% and 45%, respectively, of the equivalent empirical change in the period. As already noted, we do not want to encourage such empirical interpretations too

show the same graph but restricted to the lower half of percentiles in Figure 1.7b.

There are a number of reasons why one might consider the effect sizes we present conservative. First, the elasticity of substitution between the two skill types may be smaller in practice, leading to larger price effects: the elasticity has been measured in the previous literature using data on college versus non-college educated labor. However, that categorization is a noisy measure of the underlying skills that our theory predicts is relevant. This would lead to an overestimation of the elasticity in a typical regression methodology (e.g. in that of Katz and Murphy (1992)) due to attenuation bias, reducing the price effect (which goes to zero as the elasticity goes to infinity). Second, we have not included leisure, which works as an amplifying mechanism (cf. Guvenen, Kuruscu, and Ozkan (2014)). Third, our view of human capital is a very limited one, because we only focus on time in formal schooling. The same incentives would however also affect learning during the life-cycle, making the overall impact much larger. In addition, in this paper we are focusing on the part of the labor wedge originating exclusively from income taxation. There exist other sources for the labor wedge, in particular discrimination. Since this is outside the current model, we will postpone a detailed discussion of this to the concluding Section 1.7. Finally, potentially also the share of output the model attributes to manual skills, γ, is driving out results. However, we know little about it's empirical counterpart - this becomes a suggestion for further research. To investigate the importance of the manual skills share for our results, in the next subsection we will conduct a formal sensitivity analysis of the respective parameter, γ. As will become clear, sensitivity is relatively small. This is reassuring, as it implies that our results are relatively robust to changes in γ.

1.6.4 Sensitivity Analysis

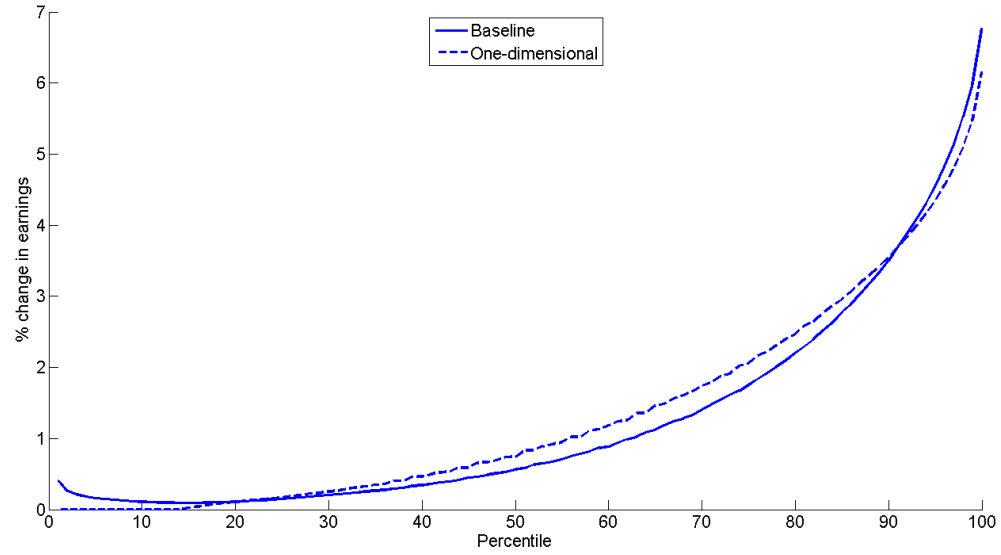
The main moment of which we are uncertain is the one informing γ, the share of output that is contributed by non-learnable skills. At the same time, this parameter is obviously crucial in assessing the importance of our mechanism: in the absence of non-learnable skills output, the model collapses to a uniform human capital model. To make this clear, we re-calibrate the model setting the moment for γ to zero, which results in $\gamma = 0$ (and slight changes to some of the other parameters). Figures 1.7a and 1.7b also show the results in this case (labeled 'one-dimensional'). While the result is similar for overall inequality, polarization has disappeared. The effect on inequality within the bottom half of the population is now much more straight-forward.[11]

much.

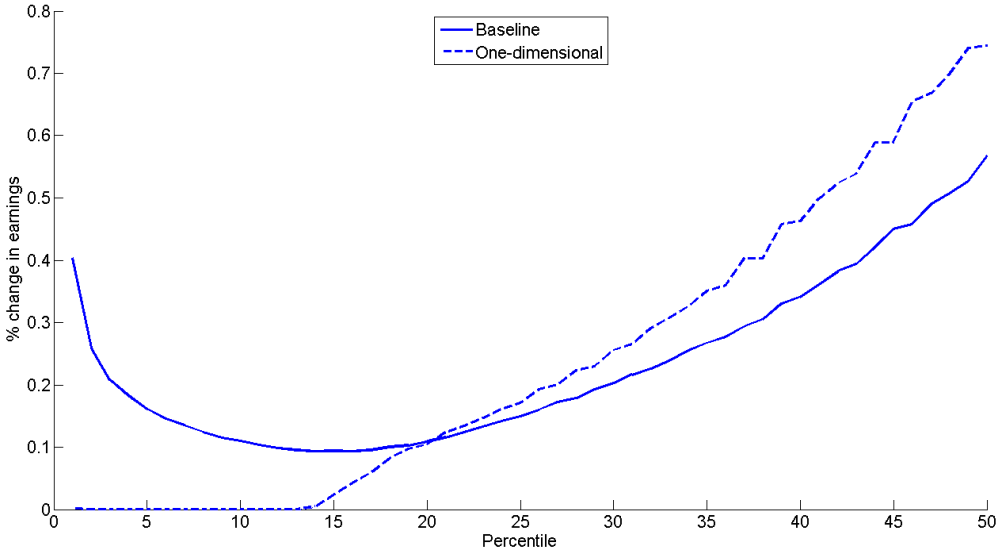
[11]The 'one-dimensional' graph in Figure 1.7b appears to display a kink that is not actually there: investment in education is always non-zero due to an Inada condition in human capital formation. The visual effect arises because levels of the human capital distribution have been compressed to a percentile

Figure 1.7: Relative earnings change under counter-factual reform

(a) Full distribution



(b) Lower half of the distribution



We provide a more formal analysis of the sensitivity of γ in the remainder of this Section. Our parameters can be interpreted as estimates of an indirect inference procedure: They are the result of minimizing the distance between the data moments described in Table 1.4, the vector of which we will now call $\hat{s}$, and the model moments that we will call $s(\theta)$ (where θ is the vector of parameters). Defining $\hat{g} = \hat{s} - s(\theta)$, we then used θ to minimize $\hat{g}' I \hat{g}$ (where I is the identity matrix that we use as weights) and reported the argmin $\hat{\theta}$ of our problem in Table 1.4.

Andrews, Gentzkow, and Shapiro (2017) establish a methodology for measuring the sensitivity of parameter estimates to estimation moments. They suggest reporting an estimate of the matrix $\Lambda = -(G'WG)^{-1}G'W$, where G is the Jacobian of the probability limit of $\hat{g}$ at the true parameter values θ_0, and W is the weighing matrix (the identity matrix in our case). The advantage of their method is that it is computationally simple to find a point estimate of G, and therefore Λ: because our objective vector $\hat{g}$ is additive and only $s(\theta)$ depends on the parameters, we can simply calculate the numeric Jacobian matrix S of our model moments $s(\theta)$ at the estimated parameter value $\hat{\theta}$. In short, we have that our *sensitivity* estimate is given by $\Lambda = S^{-1}$.

How should these sensitivity estimates be interpreted? Entry λ_{ij} of Λ tells us, roughly, how large the local impact of a change in data moment j is on parameter i. It can be used to calculate the asymptotic bias in our estimates associated with an alternative hypothesis on the data moments, as long as the alternative is sufficiently close to the data moments we report. More straightforwardly, it can be used to verbally discuss the sensitivity of our estimates to the data moments. That is a particularly appealing feature in light of the uncertainty around some of the data moments that we report above. Because a unit change in the data moments is not always easy to interpret, we instead opt to report results relevant to a 1% change of each data moment. This is achieved by multiplying λ_{ij} by a percent of data moment j. The results are in Table 1.5.

Two parameters, σ and ρ, are only sensitive to the one moment on which they depend by a closed-form relation (the latter's sensitivity measure is zero in the table due to round-off). ψ_s takes on larger values, and so is generally more responsive in level terms. ψ_m and ψ_s react most heavily to moments that describe the distribution of earnings and schooling. As we would hope, parameters describing learning ability and the formation of human capital indeed react most strongly to those moments that describe the distribution of schooling. The parameter γ reacts strongly to the 5th moment, the variance of schooling, which clearly plays an important role in the determination of the model's parameters.

As discussed above, we have very little information about the 'non-learnable share of

scale.

Table 1.5: Sensitivity Analysis

Moment nr.	1	2	3	4	5	6	7	8
σ	-0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ψ_m	0.00	-0.01	-0.03	-0.08	0.51	-0.00	-0.00	-0.01
ψ_s	0.00	50.95	335.41	-566.61	1901.24	-23.21	-2.41	129.65
μ_s	0.00	0.05	-0.02	0.29	4.45	0.01	0.00	0.02
σ_s	0.00	-0.00	0.00	-0.03	-0.19	-0.00	-0.00	-0.00
β	0.00	0.02	0.00	0.05	1.60	0.00	0.00	0.01
ρ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
γ	0.00	0.01	0.03	0.02	-0.20	0.00	0.00	0.02

Note: Model moments are 1 – Elasticity of inter-temporal substitution, 2 – Earnings variance at start of working life versus overall, 3 – Gini coefficient of gross earnings, 4 – Average share of working age spent in school, 5 – Variance of share in school, 6 – Share with zero education after age 18, 7 – Elasticity of substitution in production, 8 – Manual human capital share of output. See also Table 1.4.

output', γ, which is the eighth and last data moment in Table 1.5 above. It turns out that this parameter does play some role in the determination of γ, albeit not a large one. That there is some sensitivity is quite in line with our expectation, given the analysis included above where we set $\gamma = 0$. The fact that the sensitivity is not extremely large is reassuring, since it implies that our results would relatively little if our target of γ was somewhat off. We remain with the conclusion that the importance of manual skills in the overall economy is an important determinant of the strength of our mechanism, but that we unfortunately know little about it.

1.7 Conclusion

This paper has analyzed the effect of tax incentives on cognitive skills, in a model where (learnable) cognitive and (non-learnable) manual skills jointly produce earnings. It has also attempted to argue why this is a relevant view of the labor market, combining general equilibrium elements from the literature on skill-biased technological change and incentive elements from the literature on human capital formation. In doing so, it has provided an alternative mechanism through which labor market polarization may arise.

In the paper we focus exclusively on the part of the labor wedge originating from taxation. An important additional source of the labor wedge originates from discrimination. Over the second half of the 20$^{\text{th}}$ century (labor market) discrimination against women and non-white groups arguably decreased a lot. There is growing evidence that the decline

in discrimination has been quantitatively important for US macroeconomic outcomes. Dwyer (2013) provides evidence that polarization in employment has been driven to a substantial part by women increasingly entering the labor market, primarily in the tails of the distribution. Hsieh et al. (2016) estimate that about 25% of US output per capita growth between 1960 and 2010 can be attributed to an improved allocation of talent due declines in discrimination in the labor market and in access to education. Decreasing the price of education for a substantial share of the working population would have a similar effect as the decline in tax progressivity, by increasing the relative payoff of spending time in school. Similarly, if declines in discrimination take place in the form of 'breaking the glass ceiling', they might over-proportionally improve labor market outcomes for high-earning women, again resembling declines in tax progressivity. Potentially, these results therefore imply that the decline of progressivity of the effective labor wedge has been a lot larger than the decline in the explicit tax wedge. In this case, our results present a definitive lower bound on the supply-side polarization channel discussed here.

Future research may lead in a number of directions. First, fundamental questions on our model of the labor market remain of interest. For example, credibly exogenous variation in skill levels might illuminate the prices paid for different levels (or bundles) of skills. Second, further research into the distributional effects of reduced discrimination against minority groups in the labor market seems warranted. Finally, while the emphasis in this paper has been on positive implications, one might ask what optimal tax and education policies look like in a model like ours. In the presence of general equilibrium effects, tax disincentives to the formation of human capital are more harmful than is traditionally assumed, likely warranting less progressive tax schedules.

1.A PCA Results

The table below displays the full PCA results. Each column represents the correlation between a component and the original variables. The table begins with the component that explains the largest share of variance, then the second largest, and so forth.

Data	0.49	-0.31	0.25	0.51	0.49	0.12	0.13	0.31	0.16	0.42	0.60	0.62	0.49	-0.31	0.77	-0.06	0.81	0.70
People	0.45	-0.27	0.34	0.39	0.56	0.47	0.15	0.05	0.19	0.20	0.57	0.68	0.70	0.32	0.70	0.33	0.46	0.36
Things	-0.34	0.12	-0.49	-0.15	-0.38	-0.42	-0.15	0.30	0.56	-0.19	-0.01	-0.45	-0.55	0.11	-0.29	0.09	0.47	0.45
GED Reasoning	0.47	-0.36	0.22	0.50	0.62	0.27	0.09	0.33	0.16	0.38	0.64	0.68	0.52	0.01	0.66	-0.19	0.86	0.77
GED Mathematical	0.46	-0.39	0.22	0.49	0.45	0.04	0.06	0.43	-0.02	0.55	0.56	0.62	0.45	-0.01	0.64	-0.35	0.83	0.72
GED Language	0.48	-0.38	0.27	0.54	0.61	0.34	0.04	0.43	0.14	0.32	0.65	0.69	0.65	0.04	0.67	-0.22	0.75	0.70
Specific Vocational Preparation	0.36	-0.22	0.04	0.46	0.36	0.03	0.17	0.24	0.29	0.31	0.46	0.61	0.31	-0.20	0.58	-0.13	0.88	0.75
Intelligence	0.51	-0.44	0.26	0.54	0.60	0.25	0.05	0.36	-0.02	0.39	0.72	0.69	0.52	-0.09	0.54	0.02	0.80	0.75
Verbal	0.51	-0.49	0.37	0.61	0.55	0.34	0.09	0.19	0.00	0.37	0.76	0.68	0.67	-0.15	0.60	-0.01	0.72	0.69
Numerical	0.43	-0.62	0.31	0.38	0.59	0.02	-0.06	0.38	0.11	0.59	0.58	0.53	0.36	-0.02	0.53	-0.22	0.71	0.72
Spatial	0.05	0.41	-0.59	0.07	0.13	-0.63	0.21	0.12	0.19	0.12	-0.15	0.02	-0.19	0.16	0.14	-0.03	0.50	0.56
Form Perception	0.05	-0.09	-0.32	0.49	-0.04	-0.09	-0.06	0.34	0.42	0.30	0.04	-0.11	-0.00	-0.03	-0.04	0.21	0.63	0.76
Clerical Perception	0.53	-0.51	0.60	0.42	0.38	0.45	-0.45	0.34	0.07	0.41	0.73	0.40	0.41	-0.09	0.41	-0.04	0.47	0.56
Motor Coordination	-0.44	-0.07	-0.47	-0.37	-0.55	0.01	-0.35	0.05	0.25	-0.14	0.08	-0.51	-0.24	0.08	-0.42	0.28	0.27	0.49
Finger Dexterity	-0.63	-0.14	-0.31	-0.00	-0.37	-0.05	-0.40	0.15	0.39	-0.03	0.04	-0.48	-0.15	0.08	-0.36	0.22	0.40	0.57
Manual Dexterity	-0.57	0.21	-0.54	-0.38	-0.45	-0.48	-0.13	-0.19	0.22	-0.47	-0.57	-0.70	-0.36	0.24	-0.53	0.11	0.15	0.15
Eye-Hand-Foot Coordination	-0.14	0.44	-0.51	-0.43	-0.30	-0.26	0.48	-0.20	-0.22	-0.69	-0.38	-0.17	-0.61	0.46	0.09	0.20	-0.16	0.01
Color Discrimination	-0.28	0.46	-0.05	-0.25	-0.09	0.16	0.51	0.60	0.39	0.00	-0.24	-0.27	-0.14	-0.00	-0.02	0.28	0.20	0.43

Chapter 2

The WTO Government Procurement Agreement as a Commitment Device: A First Economic Appraisal

This Chapter is joint work with Bernard Hoekman. It has been previously published as CEPR Discussion Paper No. 13266.[1]

Abstract

This paper presents novel reduced form evidence on the association between international trade agreements that include disciplines on public procurement practices – reflected in the WTO Agreement on Government Procurement (GPA) and preferential trade agreements (PTAs) – and public sector imports following the 2008 financial crisis. The results are suggestive of such international disciplines acting as an effective commitment device: GPA membership is associated with a significantly higher import share following the 2008 financial crisis than is observed for countries that are not members. We also find evidence that the GPA and PTAs that cover public procurement are partial substitutes.

[1]We are grateful to Doug Nelson, Peter Egger, Matteo Fiorini, Joe Francois, Dennis Quinn, Alan Winters and participants in the workshop Policies, Politics and Corporate Dimensions of Globalization, May 21-22, 2018 for helpful comments and suggestions. We are especially indebted to Anirudh Shingal for sharing data. The project leading to this paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770680.

2.1 Introduction

Public procurement constitutes a large market in all economies, comprising more than 10 percent of GDP in many countries (Djankov, Islam and Saliola, 2016). As has been documented extensively, governments generally source a smaller share of the goods and services they purchase from foreign countries and foreign firms than the private sector does (Breton and Salmon, 1995; Evenett and Hoekman, 2005; Shingal, 2015). This 'home bias' provides an incentive for countries to negotiate disciplines on public procurement in international trade agreements. Examples include the plurilateral WTO Agreement on Government Procurement (GPA), as well as, increasingly, bilateral or regional preferential trade agreements (PTAs). The inclusion of government procurement practices in trade agreements is relatively recent. Government procurement was excluded from the original General Agreement on Tariffs and Trade (GATT) in 1947. It was not until the late 1970s that the first iteration of the GPA was negotiated. This covered just a subset of GATT contracting parties, something that continues to be the case under the WTO. Inclusion of public procurement in PTAs is even more recent. Until the early 1990s, most PTAs did not cover procurement.

In this paper we assess the relationship between participation in international trade agreements that include government procurement disciplines and public sector imports following the 2008 financial crisis. Our focus is on the WTO GPA and the subset of PTAs that include public procurement provisions. We investigate whether countries that have made commitments on public procurement practices maintain higher levels of public sector openness after the 2008 financial crisis than countries that have not done so. Our hypothesis is that governments are likely to have greater incentives to steer public funds towards domestic economic operators following a major shock to aggregate demand. An implication is that we expect to observe a difference in the public procurement behavior of countries depending on whether they have signed binding (and enforceable) agreements not to discriminate against foreign companies when procuring products.

We find a statistically significant difference in the behavior of GPA members versus non-GPA signatories, with GPA countries maintaining higher levels of public sector openness in the post-crisis period. We also find indicative evidence that PTAs with procurement disciplines may partially act as a substitute for the GPA. To our knowledge this is the first empirical analysis of the role of the GPA as a commitment device and the first to consider the interaction between GPA membership and PTA-based mechanisms to discipline public procurement policies.

The existing policy literature on the relationship between trade agreements and public procurement focuses primarily on the extent which the GPA and PTAs provisions increase

access to procurement markets by evaluating the legal texts (coverage) of trade agreements. The limited empirical literature tends to explore whether negotiated commitments to reduce discrimination against foreign products lead to greater foreign sourcing. The basic finding is that although agreements have gradually increased the coverage of public procurement they do not appear to increase foreign sourcing (e.g.,Rickard and Kono, 2014).[2] A growing literature on international trade policy stresses the potential role of international agreements in curbing policy uncertainty and thereby facilitating international trade, especially in periods of economic distress (Carballo, Handley and Limão, 2018).[3] Curbing international policy uncertainty provides an alternative economic impact channel for the GPA.

The 2008 financial crisis constituted a major exogenous shock that permits analysis of the role, if any, played by inclusion of procurement-related disciplines in trade agreements. The shock generated pressures on governments to use fiscal policy to support domestic activity, including through government demand, and therefore can be expected to have increased the incentives for government agencies to allocate procurement expenditures to local firms to support local employment.[4] From the point of view of potential foreign bidders, this shift in home-bias potentially acts as an increase in uncertainty regarding the prospects of success in bidding for contracts. The post-crisis period therefore provides a natural experiment to assess whether countries that are members of the GPA and (or) that have signed PTAs with public procurement provisions (PP-PTAs) display different behavior from countries that are unconstrained by such agreements. More specifically, the crisis allows for a test of the commitment function of trade agreements.

We use data on aggregate public import shares constructed from the World Input Output database (WIOD) for the 2000-2014 period. The resulting public import penetration (PIP)

[2]One possible reason for this may be that procurement agreements do little to change actual applied policies but are used primarily to bind or 'lock-in' status quo policies. If so, they are unlikely to have substantial effects in increasing effective market access and thus drive changes in sourcing patterns. More generally, the impact of the GPA has been difficult to identify empirically because membership has largely been invariant over the time period for which data are available. Moreover, many of the countries that joined the GPA in the 1990s and 2000s were European, and signing the GPA was a corollary of accession to the European Union. This makes it very difficult if not impossible to distinguish the GPA effect from the broader regime shift that occurred for the countries concerned.

[3]Francois and Martin (2004) is an early contribution highlighting the role that policy commitments in trade agreements can play in influencing investment decisions of firms by reducing uncertainty. More recently, Handley (2014) and Handley and Limão (2015, 2017) provide estimates of the effects of such uncertainty reduction.

[4]Other papers in the procurement literature have found that home bias increases in recessions – see e.g., Shingal (2015). Evenett and Shingal (2016) document a variety of measures taken by countries post-2008 to allocate more procurement expenditures to local firms.

measure is based on the public consumption category of final use in national accounts, disaggregated at the country-industry level. While the aggregate nature of the data imposes limitations, e.g. some elements of the flows are based on imputations, the data are internationally consistent and allow an initial investigation whether binding commitments on public procurement policies appear to be effective. We run a set of cross-country panel regressions with country fixed effects and a variety of controls. The results suggest that GPA membership is associated with less change in average PIP levels post-crisis, and significantly higher PIP levels. This finding is robust to controlling for the number and depth of PP-PTAs. While the results are not driven by exogenous variation, limiting the causal interpretability of the results, they lend some credence to the interpretation that the GPA was instrumental in sustaining levels of public sector openness. Given that comparable evidence is largely absent in the literature, we view the results as a useful first step in assessing the role that trade agreements can play as a policy commitment (uncertainty reduction) mechanism.

The plan of the paper is as follows. In Section 2 we discuss the extant literature on international policy agreements and the effects of the GPA. Section 3 describes the data used in the empirical analysis. Section 4 presents the methodology and presents the results. Section 5 concludes.

2.2 Background and Literature Review

In this Section, we briefly review some of the salient literature on trade policy, trade agreements and public procurement, the WTO GPA and the evolution of PP-PTAs.

2.2.1 Trade policy and trade agreements

Three broad rationales for trade agreements have been developed in the literature. One centers on market access as a mechanism to reduce terms of trade externalities created by national trade policies that are set non-cooperatively. The basic argument is that countries seek to negotiate away the negative terms-of-trade spillovers generated by the imposition of trade restrictions in partner countries (Bagwell and Staiger, 2002). Another strand of literature argues that trade agreements offer a mechanism (independent of terms-of-trade considerations) to governments that want to adopt policies that are not politically feasible to implement or maintain as a result of time inconsistency or credibility problems (e.g., Maggi and Rodriguez-Clare 1998). By committing to rules that constrain policy choice, governments can make policy reforms more credible. A third perspective stresses political economy drivers (e.g., Ethier, 2007) and the premise that governments seek to maximize political support. Assuming that governments put greater weight on prospective losses for groups in society than on the expected gains from liberalization (Deardorff, 1987),

governments have incentives to impose or maintain protection because this raises the incomes of the groups from which they derive political support. If foreign governments can be induced to liberalize, however, in the context of a trade agreement, that provides a direct gain for existing exporters and shift the balance of domestic political support towards liberalization.[5]

A corollary of all these potential rationales for a trade agreement is that they reduce policy uncertainty for firms and traders. This is because agreements entail binding commitments on current policies and future trade policy changes. If exporting entails sunk investment costs upon market entry, the exporting decision will have a dynamic component and the expected return will be a function of expected future trade policy. Even a commitment not to exceed a certain level of protection (as is the case with tariff bindings) that is well above actually applied protection has value by limiting the prospects of confronting high trade barriers in a future state of the world. Thus, trade agreements establish an upper bound on the downside risk confronted by traders and investors (Francois and Martin, 2004). By reducing the maximum *potential* level of protectionism, international agreements play a role in removing the option value of waiting to resolve policy uncertainty, which may trigger investment in tradable activities even in the absence of actual changes in policy.

The role of international agreements as a mechanism to lower policy uncertainty becomes especially salient during periods of economic distress, when governments are more prone to consider discrimination against foreign products. Recent research by Handley and Limão has demonstrated the empirical salience of this dimension of trade agreements. Investigating the impact of accession to the WTO by Australia in 1996, Handley (2014) finds that the accompanying reduction in trade policy uncertainty substantially reduced barriers to entry and that exporter product variety growth would have been 7% lower in the absence of the WTO tariff bindings. Handley and Limão (2015) investigate the effect of Portugal becoming a member of the European Community (EC) in a structural dynamic model with sunk export costs. Their model attributes a large fraction of the observed growth in Portuguese exporting firms to the removal of future policy uncertainty associated with adoption of the EC common commercial policy. Handley and Limão (2017) find large effects also for the US, following the accession by China to the WTO in 2001, taking into account general equilibrium effects on prices. Overall, an emerging body of evidence suggests that the effects of policy uncertainty on firms'– investment decisions can be large, that international agreements can play an important role in reducing trade policy uncertainty, and that this is associated with substantial economic gains for the participating economies.

[5]The various theories that have been developed to explain the role of trade agreements are surveyed and summarized in Grossman (2016).

2.2.2 The GPA and PTAs with procurement provisions

Public procurement was excluded from the GATT and older vintage PTAs because it was regarded as a state activity as opposed to a commercial one. Procurement gradually came to be covered by trade agreements because of the magnitude of the associated markets. These motivated pursuit of reciprocal reductions in the incidence of explicit discriminatory policies that impede market access opportunities for foreign firms. The first version of the GATT GPA entered into force in 1981. It was revised several times subsequently to expand its coverage. At the time of writing, there are 19 parties to the agreement, counting the EU-28 as one, so that the GPA covers 47 WTO members.[6] The agreement is a so-called plurilateral agreement in that it binds only signatories. The benefits of membership are restricted to signatories (Evenett and Hoekman, 2005).

The main discipline imposed by the GPA is that covered government entities are prohibited from discriminating against or between foreign products and firms (bidders) in the award of procurement contracts that exceed certain value thresholds. The obligation extends not only to imports but also to subsidiaries of locally established foreign firms. As foreign direct investment (FDI) is an important channel for firms to contest procurement markets, this is an important feature of the GPA. The agreement has extensive provisions aimed at ensuring that firms can become aware of procurement opportunities, that the process is transparent, and that competitive procurement methods be used in selecting and awarding contracts. There are numerous provisions that aim to realize these objectives including that notices of intended or planned procurement are published, minimum time periods for bids, economic and technical requirements, terms of payment, etc.

Very few developing countries have joined the GPA, reflecting concerns that thwe GPA impedes the ability to pursue industrial policy objectives and that national firms only have limited ability to contest foreign procurement markets. These two factors significantly reduce the incentive to engage in reciprocal negotiations to open procurement markets. The lack of interest by many countries to join the GPA has led incumbent members to pursue efforts aimed at extending procurement disciplines through the negotiation of PTAs. Developing economies may be more willing to sign a PTA that includes procurement liberalization given that concessions may be offered in other areas. This is not possible in the GPA context given that the GPA deals only with public procurement.

[6]GPA membership spans Armenia, Austria, Belgium, Bulgaria, Canada, Chinese Taipei, Croatia, Cyprus, Czech Republic, Denmark, Estonia, European Union, Finland, France, Germany, Greece, Hong Kong, China, Hungary, Iceland, Ireland, Israel, Japan, Korea, Rep., Latvia, Liechtenstein, Lithuania, Luxemburg, Malta, Moldova, Rep., Montenegro, Netherlands, Netherlands for Aruba, New Zealand, Norway, Poland, Portugal, Romania, Slovak Republic, Slovenia, Singapore, Spain, Sweden, Switzerland, Ukraine, United Kingdom, USA.

Several efforts have been made to classify the scope and coverage of procurement in PTAs.[7] What follows briefly describes a recent effort by Shingal, Ereshchenko and Mattoo (2018) to assess the coverage of procurement provisions in PTAs. Of a total of 242 PTAs currently in force that include at least one non-GPA signatory, 127 (52 percent) include language on public procurement (Annex Table A1 lists the PTAs in the dataset and indicates which PTAs include provisions on procurement). Of these 127 PTAs, 64 (50 percent) specify in some detail what types of procurement are covered (Table 1). Most PP-PTAs are modelled to a greater or lesser extent on the GPA (Anderson et al. 2011). Few go beyond the GPA in terms of rules or coverage although some do. Thus, a little less than one-sixth include threshold values that are lower than those applying in the GPA, implying that applicable rules of the PTA apply to a larger share of calls for tender (Shingal et al. 2018).

Table 2.1: Depth of procurement commitments in PTAs with at least one non-GPA signatory

Criterion	Frequency (%)
Government procurement coverage is detailed in the agreement	49.6
The agreement covers only central government entities	8.5
Threshold values for purchases of goods are lower than in the GPA	14.7
Procurement provisions are enforceable (incl. via domestic review)	37.2

Note: Sample comprises 127 PTAs. See Annex Table A2.
Source: Shingal et al. (2018).

Of particular importance from the perspective of credibility of commitments is whether provisions are enforceable. A total of 48 of the 127 PTAs have a hard law dimension in the sense that at least some provisions can be invoked in formal dispute settlement procedures and through domestic review ('bid challenge') mechanisms that permit firms to contest ongoing procurement tenders and awards. These types of requirements are also a key feature of the GPA.[8] Provisions calling for tenders be published, that bids are opened in public, that procuring entities must award contracts to the lowest bid that satisfies the technical criteria, and so forth, are much less relevant to firms if there is no effective

[7]Examples include Anderson et al. (2011) and Ueno (2013), who conclude that non-GPA countries accept a level of procurement market access commitments in PTAs that are very similar to those that are contained in the GPA.

[8]Most of the PP-PTAs that include binding (i.e., enforceable) procurement provisions include one or more OECD member countries (see Annex Table A2). But there are also South-South PTAs involving Central American states, Chile, Colombia and Peru. In addition to PTAs with other OECD member countries, Chile, for example, has PP-PTAs with Costa Rica, Honduras, Guatemala, Nicaragua, and El Salvador. There are no PTAs with serious coverage of procurement in Africa, the Middle East, or South Asia.

recourse if entities do not follow the rules. Domestic review mechanisms are in practice the primary if not sole enforcement mechanism for firms, as State-to-State dispute resolution is slow and does not offer a prospect of meaningful remedies, as these are prospective in nature. Domestic review generally provides the possibility of compensation for tendering costs, damages and legal fees.[9]

2.2.3 Related literature

Evenett and Shingal (2006) and Shingal (2015) have used data reported by the few countries that provide statistics to the WTO on the national breakdown of winning tenders on contracts that are covered by the GPA, focusing on Japan and Switzerland. Evenett and Shingal (2006) conclude that a smaller share of contracts above the value thresholds established by Japan was awarded to foreign suppliers in 1998-99 upon signing the GPA than in 1990-91, prior to joining. Shingal (2015) analyzes the determinants of procurement sourcing over time in these two countries, controlling for factors that may affect sourcing from foreign firms such as the state of the business cycle and overall trade policy trends and trade costs. He finds that GPA membership has no independent effect on sourcing behavior. The same conclusion emerges from an analysis of the extension of the GPA in 1996 to include services procurement. Again using data reported to the WTO by Japan and Switzerland, Shingal (2011) finds that the share of services contracts awarded to foreign suppliers declined over time for both countries. Similarly, Rickard and Kono (2014), focusing on overall import penetration, conclude that GPA membership has no impact on the ratio of imports to government demand.

In contrast, Chen and Whalley (2011) find that the GPA has a positive impact on trade both among members and with third parties. However, they rely on self-reported public procurement trade notifications to the GPA committee which are of low quality for most countries. Tas, Dawar, Holmes and Togan (2018) focus on EU procurement. Using very detailed transaction-level data on procurement awards from the EU Tenders Electronic Database, they assess the effects of the GPA on procurement market openness of EU countries. They conclude that the GPA increases the probability of a contract being awarded to a foreign firm. They also find that the GPA reduces the risk of corruption by decreasing the number of contests with single bidders and the number of wins by a single firm. Identifying the independent effect of the GPA in the context of the EU is difficult given that EU procurement law and policy is both broader and deeper than the GPA.

There is even less empirical research on PTAs than there is on the GPA. Rickard and Kono (2014) assess the effects of 43 PTAs that include procurement, focusing on overall import

[9]Domestic review mechanisms take various forms. See World Bank (2016) for an overview of national public procurement regimes.

penetration and find that they have no impact on the ratio of imports to government demand. Using data from the US Federal Procurement Data System set for 1996–2010, Fronk (2015) finds a statistically significant positive effect of US agreements including procurement. Using a gravity regression framework and focusing on agreements negotiated by the US (including the GPA) that require national treatment of foreign bidders, Fronk (2015) finds that these result in a 150 percent rise in the number of contracts won annually by foreign bidders for tenders that exceed the value threshold of the relevant agreement. However, this is only equivalent to an additional 135 contracts – reflecting the fact that the overwhelming majority of contracts (some 98 percent) is awarded to US firms. Thus, there is an effect, but because the baseline level of foreign awards is small, the magnitude of the impact is also quite small. Because US data on nationality of winning bidders only starts in the mid-1990s, this analysis cannot consider the fact that the countries that mostly win procurement bids in the US (Canada, EU, Japan) have been members of the GPA from the start (1981) and thus that much of the procurement that is analyzed was already subject to disciplines for a long period of time. It is therefore not necessarily the case that the positive sourcing effect attributed to the agreements is in fact due to them as opposed to other factors.

Gourdon and Messent (2017) combine annual bilateral import data from UN Comtrade covering 74 countries, 44 of which had signed PTAs with procurement provisions with at least one other country in their sample, and the EU Tenders Electronic Daily database for the period between 2009 and 2014 to analyze the determinants of EU procurement sourcing. They find that the GPA increases the probability that foreign firms from a GPA member win contracts in the EU. Gourdon and Messent (2017) also document that restrictive FDI policies may undercut the impact of the GPA in expanding foreign sourcing. They conclude that the GPA has an effect in reducing home bias in procurement awards.

A small number of papers use international input-output data of the type we use in our analysis. Aguiar et al. (2016) and Kutlina-Dimitrova (2017) work with GTAP 9 data on public procurement and calculate counter-factual scenarios that remove observed government home bias, finding large effects on GDP and economic welfare. Crespi and Guarascio (2017) use WIOD to calculate public procurement openness, following the definition of Messerlin and Miroudot (2012). They are concerned with measuring the effect of procurement internationalization on domestic innovative activity. Mulabdic and Rotunno (2017) similarly rely on the Messerlin and Miroudot (2012) methodology and use OECD TiVA data to estimate bilateral gravity models of private versus government sector openness. Their data comprise 5 year intervals between 1995 and 2009. They find that EU membership has strong positive effects on public procurement openness. PTAs do so as well but to a smaller extent, while the GPA has no effect on public procurement

openness.

2.3 Openness of public procurement markets, the GPA and PTAs: Descriptive evidence

The extant empirical studies of the effects of the GPA and PTAs in reducing home bias in the award of procurement contracts come to ambiguous conclusions. Studies using detailed micro data tend to be country or EU-specific. Despite its large share of GDP, comparable disaggregated data on public procurement contracts and their allocation between national and foreign bidders (suppliers) is not available on a cross-country basis.[10]

For our analysis of the impact of GPA membership and other PTAs on public sector openness during the 2008 financial crisis we make use of data from WIOD. To the best of our knowledge, our study is the first to estimate the effect of the GPA on trade using government imports of *value added*. WIOD is one of a number of initiatives that have emerged integrating and linking national IO tables across countries to provide data on international economic linkages at the industry level. One advantage of using international IO tables for analyzing public sector imports is that imports are defined consistently across countries. The 2016 release of WIOD covers 43 countries from 2000 to 2014 and includes the majority of GPA member countries as well as the major emerging economies that are not GPA members.[11] The set of countries included cover 85-90 percent of world GDP over the sample period (Table 2). In earlier years GPA members make up most of the sample GDP. This share falls to 68 percent in 2014, due to strong economic growth of emerging economies that are not GPA members.

WIOD data on the country-industry decomposed share of imported products in total government consumption are constructed on the basis of overall imports of different types of goods as reflected in UN Comtrade statistics. Thus, they do not reflect actual reported imports by governments as such data are not collected in national accounts statistics. In the absence of detailed comparable data on actual procurement, WIOD offers a consistent

[10]Some countries report detailed contract level data on public procurement awards, e.g., the European Union, Peru, Brazil, South Korea, Turkey and the US. With the exception of the EU and US, these countries are not GPA members. Papers using these data to assess the determinants and economic effects of contract awards include Kutlina-Dimitrova and Lakatos (2016), Hebous and Zimmerman (2016), Ferraz et al. (2015) and Lee (2017). For example, Onur, Ozcan and Tas (2012) analyze 90,000 government procurement tenders held in Turkey during the 2004–06 period. They find that the number of bidders significantly and negatively impacts on the prices paid and that opening tenders to foreign participation further lowers prices paid.

[11]Details on the construction of WIOD can be found in Timmer et al. (2015). We chose WIOD over the OECD ICIO database as the latter only has data through 2011.

Table 2.2: Share of included countries in global GDP

	Share of Sample in World Output		Share of GPA in Sample Output	
year	Gross output	Value added	Gross output	Value added
2000	0.14	0.86	0.09	0.91
2001	0.14	0.86	0.09	0.91
2002	0.14	0.86	0.09	0.91
2003	0.14	0.86	0.09	0.91
2004	0.15	0.85	0.09	0.91
2005	0.17	0.83	0.10	0.90
2006	0.18	0.82	0.11	0.89
2007	0.21	0.79	0.12	0.88
2008	0.23	0.77	0.13	0.87
2009	0.24	0.76	0.12	0.88
2010	0.27	0.73	0.13	0.87
2011	0.29	0.71	0.14	0.86
2012	0.30	0.70	0.15	0.85
2013	0.32	0.68	0.15	0.85
2014	0.32	0.68	0.14	0.86

Source: Own calculations based on WIOD 2016 database.

and comparable set of proxies for government imports. Even though the specific figure for the share of government consumption in total national imports is unlikely to accurately reflect actual procurement patterns in the country concerned, because the WIOD data are constructed on a consistent basis in the same way for all countries, they nonetheless permit an assessment of differences across countries in government imports at a point in time as well as trends over time.

To motivate the subsequent empirical analysis, Figure 1 plots the average PIP for GPA member and non-member countries over time. Two patterns stand out: First, in the pre-2008 period, public import shares of GPA members and non-members follow similar trends. For both groups, public import shares are rising. Second, PIP evolves differently between the two groups following the 2008 financial crisis. While initially PIP experiences a substantial drop in 2009 for both groups, it recovers for GPA member countries and remains high through the end of the sample period (2014). For non-GPA member countries on the other hand, the drop in PIP is bigger, the immediate rebound of the import share is less pronounced and the average public import share declines after 2010, reversing the pre-crisis positive trend. Furthermore, while non-member–s average PIP was higher pre-crisis than for GPA members this pattern is reversed after 2008.

Figure 2.1: Public Import Penetration by GPA membership over time

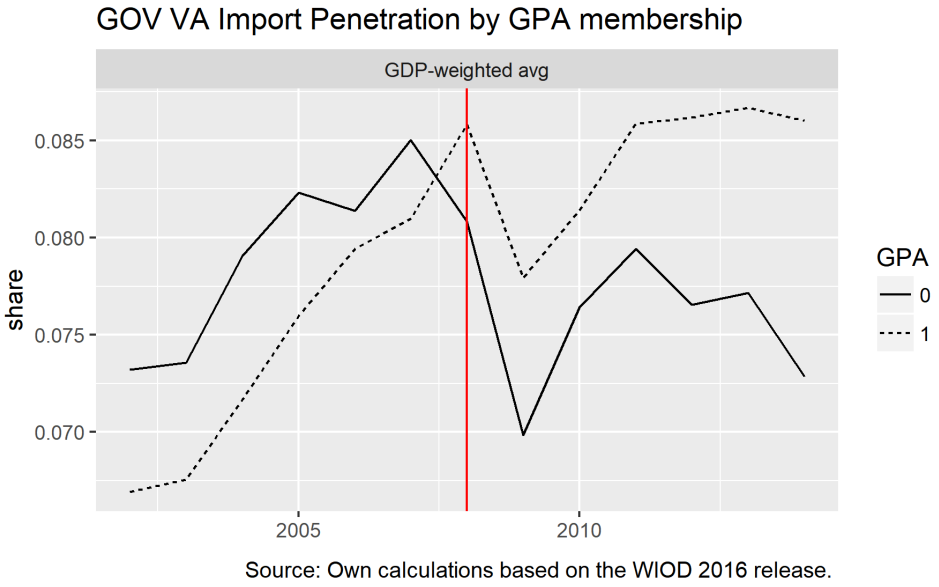


Table 3 provides more texture, reporting PIP ratios for selected countries for three time periods, 2000-02, 2007-09 and 2012-14. The countries selected are large, given a presumption that small countries will generally display higher PIPs as they have fewer opportunities to source nationally from globally efficient firms. The data indicate that the PIP ratios for the EU28 are some 10 percent higher than those of the US,[12] and both import less than the simple world average (7 percent) which is to be expected given that large economies will be better able to source domestically.

The non-GPA members included in Table 3 source more from abroad than the GPA members during the first two periods: 6.2 vs. 5.7 in the 2000-02 period. This changes over the course of the decade: In 2007-09 the PIP ratios for the two groups are roughly the same, and by 2012-14 the GPA countries have higher PIP ratios than the selected non-GPA members (6.5 percent vs. 5.6 percent). However, the selected countries in both groups source less from the rest of the world than the world average - presumably reflecting their above average size. There are large differences across countries, with Brazil only sourcing 2.9 percent from abroad, as compared to Korea at 11.6 percent in 2012-14. Some of the larger non-GPA countries appear to have been shifting steadily away from foreign sourcing since 2000. This is the case for India, Turkey and China (post-2008). On average, GPA members see a more rapid increase in foreign sourcing during the 2000-14 period, resulting in convergence over time towards the average level of 'openness' of the world as a whole, confirming the unweighted results from Figure 1.

[12]Extra-EU in Table 3 measures the import content from non-EU countries, i.e., it excludes intra-EU sourcing by the public sector.

Table 2.3: Government consumption (import penetration ratios, selected countries)

	avg. 2000-02	avg. 2007-09	avg. 2012-14	% Δ 2000-08	% Δ 2008-14
Non-GPA-members:					
Australia	6.1	5.9	5.5	-4.2	-6.7
Brazil	2.2	2.5	2.9	11.7	18.9
China	3.7	4.8	4.0	31.7	-17.0
India	4.0	5.5	5.5	35.1	-0.1
Indonesia	10.2	7.0	6.4	-31.5	-7.8
Mexico	5.8	6.1	6.6	4.5	8.3
Turkey	11.2	8.1	6.2	-27.7	-23.8
Average	6.2	5.7	5.3	-8.0	-6.8
GPA members:					
Canada	4.7	5.2	5.1	10.6	-1.8
Extra-EU	3.7	4.6	4.8	23.8	6.2
Japan	2.2	4.4	6.6	103.8	50.8
Korea	8.7	9.6	11.6	11.3	20.2
US	3.3	4.1	4.1	26.4	-1.2
Average	4.5	5.6	6.5	24.3	15.4
World	5.7	7.0	7.0	21.9	0.8
Memo: Intra-EU	3.8	4.6	4.8	22.0	2.7

Note: Averages are simple country averages.
Source: WIOD 2016 database.

Table 4 reports the results of a simple OLS regression of PIP by country across time, distinguishing between GPA members and non-GPA members, as follows:

$$PIP_t = \beta_0 + \beta_1 GPA + \beta_2 Trend + \beta_3 GPA \times Trend + u_t$$

where PIP is defined as imports of value added in government consumption final demand, GPA is a dummy variable equal to 1 if a country is a GPA member and Trend is a linear annual trend variable equal to 1 in the base year 2000.[13] This model is estimated using data on all countries included in WIOD (43 countries plus a residual –rest of the world– variable). GPA members show a larger share of foreign sourcing over the period covered when conditioning on GPA membership exclusively (column 1), but this is something that occurred for all countries over the time period (column 2). Column 3 reports the results including the interaction effect between GPA membership and the time trend. While non-GPA members show no signs of growth in PIP over the sample period, GPA

[13]The construction of PIP is discussed in greater detail in the next section.

members exhibit a significantly positive trend in PIP between 2000 and 2014. We take these simple regression results as suggestive that GPA membership may play a role in sustaining public sector openness.

Table 2.4: PIP trends over time by GPA membership

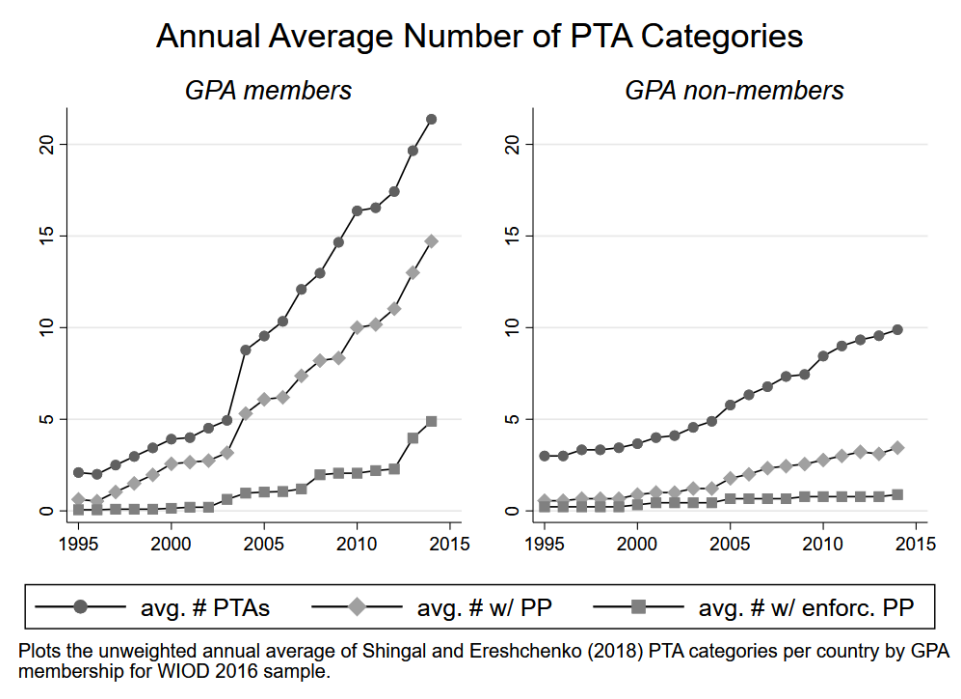
	(1)	(2)	(3)
GPA countries	0.0461***		0.0218**
	(0.0030)		(0.0071)
Trend		0.0015***	-0.0009
		(0.0004)	(0.0006)
GPA*Trend			0.0030***
			(0.0007)
Constant	0.0606***	0.0854***	0.0681***
	(0.0025)	(0.0032)	(0.0062)
Observations	660	660	660

Note: $^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$. Robust standard errors in parentheses.

Figure 2 plots the evolution of procurement coverage of PTAs over time, which we use as an additional explanatory variable in the empirical analysis. The underlying data span all agreements included in Shingal, Ereshchenko and Mattoo (2018) that have at least one participating country from the WIOD sample. Figure 2 plots the trend in the simple annual average number of three categories of PTAs (any PTA, those including text on public procurement, and those that have enforceable provisions), differentiating between GPA members and non-GPA members. Two features are worth noting. First, the growth in the number of PTAs since the 2000s is primarily driven by GPA member countries. Up to the early 2000s, the overall number of agreements is similar for member and non-member countries. Since then it has grown substantially for GPA member countries, while the increase has been much more modest for non-GPA members.[14] Second, the increase in the number of PTAs is driven by an increase in PTAs that include public procurement provisions. While the number of PTAs with enforceable procurement provisions has been growing more slowly, it has accelerated during the last decade.

[14]The pronounced jump from 2003 to 2004 is driven by the 2004 EU enlargement.

Figure 2.2: Evolution of Public Procurement Provisions in PTAs



Plots the unweighted annual average of Shingal and Ereshchenko (2018) PTA categories per country by GPA membership for WIOD 2016 sample.

2.4 Empirical analysis

Both the trends in aggregate PIP for the countries included in Figure 1 and the post-crisis differentials in openness suggested by Table 3 suggest that GPA membership may have played a role in sustaining public sector openness following the 2008 financial crisis. In this Section we use the panel structure of our data to investigate the impact of GPA membership empirically. While this does not permit a causal interpretation, the aim is to assess whether GPA membership is associated with patterns of foreign sourcing during and after the 2008 financial crisis and how GPA membership interacts with alternative sources of international procurement discipline, i.e., PP-PTAs. We also explore the robustness of the results to inclusion of a variety of alternative determinants of PIP, including country size, level of development and the general level of restrictiveness of trade and investment policies.

We regress PIP by country over the 2000-14 period on GPA membership status, the number of PTAs and the type of PTA, distinguishing between PTAs that have procurement provisions and those that are enforceable. We use a fixed effects methodology, employing country-level fixed effects. This precludes direct estimation of the effects of GPA membership on PIP, as it is collinear with the fixed effects because GPA membership is a variable that is constant during the time period considered for practically all countries in the sample. However, the approach does allow for assessment of the effect of GPA membership in the post crisis period by including an interaction effect between GPA

membership status and the crisis. This is our variable of interest.

The model we estimate is:

$$PIP_{it} = \beta_1 Crisis_t + \beta_2 GPA_i + \beta_3 Crisis_t \times GPA_i + \beta_C Controls_{it} + \alpha_i + u_{it}. \quad (2.1)$$

PIP is defined as imports of value added in the government consumption final demand category of WIOD. Following Koopman et al. (2014) country-industry value added contained in government final consumption demand $VA(FD_{gov})$ is computed as

$$VA(FD_{gov}) = \hat{V} \cdot B \cdot FD_{gov}$$

where $B = (I - A)^{-1}$ is the country-industry Leontief inverse, FD_{gov} is a country-industry $\times$ country matrix including country-industry dollar flows into the government final consumption demand of the respective country and $\hat{V}$ is a country-industry diagonal matrix with country-industry domestic value added shares on the main diagonal. Value added import shares are then obtained by summing country-industry value added imports and dividing by total value added absorbed by government final consumption demand.

$Crisis$ is a dummy variable equal to 1 for all years from 2008 onwards. GPA is a dummy variable equal to 1 for GPA members. $Controls$ include trade and FDI policies, country size and level of development. $Avg.Tariff$ is the weighted average tariff rate imposed on merchandise imports for a respective country and year. It is obtained from the World Integrated Trade Solution (WITS).[15] $FDIRI$ is an index of the degree of restrictiveness of policies towards FDI compiled by the OECD and described in Kalinova et al. (2010).[16] This is available for 1997, 2003, 2006 and on an annual basis for the post 2010 period. It is interpolated linearly for years in which the index is not reported. $No.PTAs$; $No.PTAs\,w/\,PP$; and $No.PTAs\,w/\,enforceable\,PP$ measure the number of PTAs a given country is a member of in the respective year, those that have public procurement provisions (PP); and those with PP that are binding i.e., can be enforced. These three variables are obtained from Shingal et al. (2018). All PTAs that include at least one country in our WIOD sample are counted. Enforceability refers to the PTA requiring domestic review mechanisms and permitting parties to invoke dispute settlement procedures on procurement matters. GDP and $GDP\,per\,capita$ is measured in 2010 US$ and is sourced from the World Bank Development Indicators database.

We do not include time fixed effects as this renders the $Crisis$ indicator collinear. This is problematic in our context, since the $Crisis$ indicator measures the baseline impact of the crisis for GPA non-members, against which we want to compare the $Crisis \times GPA$

[15] Available at http://wits.worldbank.org.

[16] FDIRRI data are available at https://stats.oecd.org/Index.aspx?datasetcode=FDIINDEX#.

interaction. Moreover, including time fixed effects absorbs changes in the PTA controls, since accession to a PTA is associated with a one-time jump in the associated variable. As we are interested in the relationship between PTAs and GPA membership, their number and changes therein are a key control variable.[17] We report two sets of regression results. The first focuses on the interaction between GPA membership and the number and characteristics of PTAs (Table 5). The second investigates the robustness of the results to the inclusion of the controls mentioned above (Table 6).

Column 1 in Table 5 reports results without any PTA controls. The crisis interaction with GPA membership is strongly statistically significant and positive, implying that GPA members have a higher PIP in the post crisis years compared to non-GPA countries. Columns 2 - 4 explore the effect of controlling for the number of PTAs per country, the number of PP-PTAs and the number of PP-PTAs with enforceable procurement provisions, respectively. Column 2 estimates the effect of an additional PTA, regardless of whether it includes PP provisions of any kind or not. Column 3 estimates the effect of an additional PTA with PP provisions, deep or shallow, and Column 4 estimates the effect of an additional PTA with *deep* (that is, enforceable) PP provisions on PIP.

The estimated coefficient increases in moving from Column 2 through Column 4, indicating that the number and depth of PP provisions in PTAs is indeed associated with higher government sector import shares. The significance level and magnitude of the estimated $GPA * Crisis$ interaction coefficient declines slightly but remains sizable and highly significant throughout.[18] Controlling for the overall number of PTAs and the inclusion of public procurement provisions (models (2) and (3)), yields marginally significant effects. However, once we control exclusively for PTAs that have binding public procurement disciplines (column (4)), the estimated coefficient increases substantially, both quantitatively and in statistical significance.

To sum up the results from Table 5, the $GPA * Crisis$ interaction effect is robust to the inclusion of alternative sources of international public procurement discipline (PTAs), and is almost constant quantitatively. This seems to be the case primarily for PP-PTAs that are enforceable and thus constitute more credible commitments to sustaining public sector openness. These results indicate that both GPA and PTAs are associated with high PIP

[17] If we estimate the model including time fixed effects we nonetheless obtain Crisis*GPA interaction coefficients comparable in size and significance to our main specification. Results are available on request.

[18] All results presented are based on the model presented in Equation (**??**). However, as noted previously, the GPA membership indicator GPA_i is dropped throughout due to collinearity problems. For our sample period, GPA membership is constant for almost all countries in the sample, with exception of the countries accessing the EU during this period. This implies that GPA membership is not identified separately from the country fixed effect α_i.

Table 2.5: GPA and PTA Provisions

Dependent variable:	Public Consumption Value Added Import Share			
	(1)	(2)	(3)	(4)
Crisis	−0.5585	−0.7307	−0.6545	−0.6054
	(0.3801)	(0.4082)	(0.3917)	(0.3771)
No. PTA		0.0465*		
		(0.0196)		
No. PTA w/ PP			0.0638*	
			(0.0295)	
No. PTA, enforc. PP				0.1777***
				(0.0475)
Crisis*GPA	1.6694***	1.3886**	1.3660**	1.3437**
	(0.4515)	(0.4747)	(0.4663)	(0.4398)
Country FE	Yes	Yes	Yes	Yes
Observations	660	660	660	660
R^2	0.1582	0.1766	0.1731	0.1720
Adjusted R^2	0.0965	0.1148	0.1111	0.1098

Note: $^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$. Robust standard errors in parentheses.

and their effects potentially interact.

Table 6 reports results with controls for additional explanatory variables. Column 2 includes the OECD FDI Restrictiveness Index (FDIRI) and the weighted average applied tariff on merchandise imports. These are included to control for general policies towards openness. As public procurement projects are often complex and regulation-intensive establishing a foreign affiliate is a common way to compete for contracts. To control for barriers to this indirect form of imports, we therefore include the FDIRI in addition to tariff rates on direct imports. Larger and more developed countries generally feature different levels of openness. GPA member countries are on average richer than non-members in the WIOD sample, which could potentially drive our results. A second set of controls in Column 3 controls for country size and development, proxied by real GDP and real GDP per capita.

The coefficient on direct tariff barriers is negative but insignificant throughout. This is what we would expect, since the public sector effectively does not pay tariffs. The GPA*Crisis interaction slightly decreases in size, but continues to be highly significant. A similar picture arises for GDP per capita. While the effect is quantitatively small, it also carries the expected, positive sign. When including all the controls and number of PTAs, the interaction coefficient of interest declines slightly in both size and significance

Table 2.6: GPA membership: additional controls

Dependent variable:	Public Consumption Value Added Import Share				
	(1)	(2)	(3)	(4)	(5)
Crisis	−0.5585	−0.6887	−1.1117*	−0.7801	−0.8944
	(0.3801)	(0.5211)	(0.4570)	(0.5395)	(0.5412)
FDIRI		2.9005		4.1451	6.1542
		(4.1216)		(4.4874)	(4.1917)
Avg. Tariff		−0.0868		−0.0684	−0.0746
		(0.0480)		(0.0489)	(0.0487)
GDP			0.0001		0.0002
			(0.0001)		(0.0001)
GDP/capita			0.0001***		0.0001*
			(0.00004)		(0.00004)
No. PTAs				0.0616**	0.0380
				(0.0195)	(0.0209)
Crisis*GPA	1.6694***	1.8070***	1.8708***	1.3939**	1.5031**
	(0.4515)	(0.5044)	(0.4995)	(0.5356)	(0.5338)
Country FE	Yes	Yes	Yes	Yes	Yes
Observations	660	538	630	538	538
R^2	0.1582	0.1866	0.1958	0.2098	0.2327
Adjusted R^2	0.0965	0.1194	0.1339	0.1427	0.1643

Note: $^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$. Robust standard errors in parentheses.

but remains overall robust. We interpret this as an additional support for the hypothesis that international procurement disciplines matter and interact.

Overall, the results from the panel regression suggest that GPA membership during the crisis is associated with a higher public sector openness compared to non-member countries. This conclusion is robust to general measures of trade protectionism and country size and development. GPA membership continues to be significant also when including alternative measures of international procurement discipline such as the number and characteristics of PTAs. The reduced coefficient size and significance level suggest that the GPA and PP-PTAs are partial substitutes in sustaining public sector openness.

2.4.1 Placebo regression results

The results presented in the preceding sections are indicative of GPA membership playing a role in sustaining public sector openness following the 2008 crisis. However, the nature

of our question and data prevents us from relying on controlled and exogenous variation in order to draw robust causal inference. To corroborate our results, we conduct a placebo regression exercise. This involves a replication of our main specification with an alternative outcome variable that is expected on theoretical grounds not be affected by our main explanatory variable, GPA membership. If the nature of our mechanism is causal, GPA membership should not affect the alternative outcome variable. In our case, a promising candidate as placebo outcome variable arises naturally from the fact that the GPA covers public sector imports exclusively. Accordingly, private sector imports do not fall under the GPA by definition and should be unaffected by GPA membership.

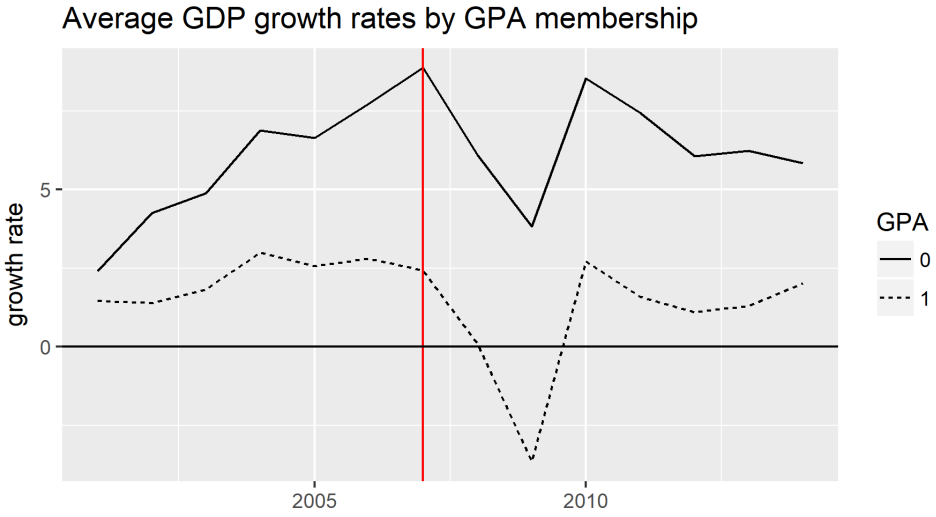
One possibility invalidating private consumption imports as placebo outcome is if there is a potential interaction between the crisis and GPA membership. As discussed previously, the GPA member countries are on average richer than non-member countries. As the crisis originated in the U.S and the EU, it may have affected developed economies in a systematically different way than emerging economies. This would invalidate our choice of private consumption imports as placebo outcome, insofar as GPA membership would mask the differential impact of the crisis because of the characteristics of the underlying economies.

To shed light on this, Figures 3 and 4 plot average GDP growth rates and private consumption import growth rates for GPA member and non-member countries, weighted by GDP. While the GDP growth rates differ between GPA members and non-members as expected, both show a comparable decline during the crisis. Similarly, average private consumption import shares seem similarly affected by the crisis. We view this as supporting the validity of our choice of private consumption import share as placebo outcome variable.

Results from the placebo regression are reported in Tables 7 and 8. Column 1 replicates the regression on the crisis and GPA interaction effect. Not controlling for other covariates, the measured impact of GPA membership is statistically significant and sizable for private consumption import shares. However, this changes once we take into account the number of PTAs in Columns 2 to 4, suggesting the GPA-crisis variable is picking up the effects of trade policy disciplines more generally. Since PTAs by definition seek to improve access to markets and reduce policy uncertainty for the private sector, the positive and significant coefficient is what we would expect. The magnitude of the estimate increases as the relevant PTAs become deeper, which is consistent with deeper PTAs providing greater security of market access conditions. In the final model that includes only the subset of PP-PTAs with enforceable procurement provisions (column 4) the crisis-GDP interaction effect becomes marginally significant again. This is not inconsistent with our hypothesis, as in this regression we omit other PTAs which target overall (private) imports. Given that PTAs with enforceable PP provisions also feature more extensive coverage and

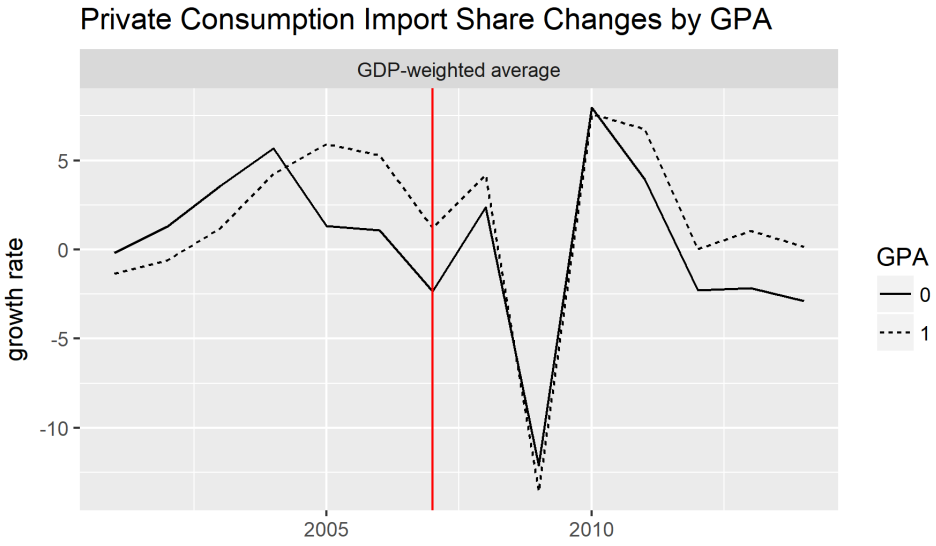
commitments in a variety of other policy areas that are salient for the private sector, this will be picked up by our measure in Column 4 and lead to a spurious correlation with our PP-PTA measure.

Figure 2.3



Note: Growth rates of real GDP in 2010 USD. Growth rates are GDP-weighted country averages. Source: World Bank national accounts data and OECD national accounts data.

Figure 2.4



Source: Own calculations based on WIOD 2016 release.

Table 8 reports results for the replication of the regression including additional controls. The covariate coefficients generally behave as expected. Coefficients of trade restrictions, both for direct imports and for foreign investment are negative and significant. GDP per capita is positively associated with private consumption imports and larger countries tend to import less, although this effect is only marginally significant. Controlling for per-capita income, which is correlated with GPA membership status, the $Crisis \times GPA$ effect

weakens and once the number of PTAs is controlled for, the effect becomes completely insignificant, as we would expect.

Table 2.7: GPA and PTA Provisions

Dependent variable:	Private Consumption	Value Added	Import Share	
	(1)	(2)	(3)	(4)
Crisis	0.2654	−0.4711	−0.1846	0.1362
	(0.4042)	(0.4485)	(0.3681)	(0.4036)
No. PTA		0.1990***		
		(0.0441)		
No. PTA w/ PP			0.2988***	
			(0.0665)	
No. PTA, enforc. PP				0.4896***
				(0.1434)
Crisis*GPA	2.1030***	0.9019	0.6811	1.2056*
	(0.5547)	(0.5628)	(0.5418)	(0.5556)
Country FE	Yes	Yes	Yes	Yes
Observations	660	660	660	660
R^2	0.2440	0.3652	0.3619	0.2817
Adjusted R^2	0.1886	0.3176	0.3141	0.2278

Note: $^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$. Robust standard errors in parentheses.

Overall, we take the results from the placebo exercise as additional confirmation for the hypothesis that the GPA, an agreement that entails binding, enforceable commitments to procure goods on a nondiscriminatory basis, served to back-stop public sector openness in the aftermath of the 2008 financial crisis.

Finally, Annex Table A-3 presents the results from a Mundlak (1978) correction regression that includes country-averages of all time-varying control variables in the estimation of a random effects model. The means of GDP, GDP per Capita and average tariffs are all significant, while the means of FDIRI and No. of PTAs are not. The latter two variables are less time-varying than the former three, which might explain this pattern. We take the results as supporting the use of fixed effects in the main specifications.

Table 2.8: Placebo GPA and additional controls

Dependent variable:	Private Consumption Value Added Import Share				
	(1)	(2)	(3)	(4)	(5)
Crisis	0.2654	−0.6859	−0.4129	−1.0055	−0.9543
	(0.4042)	(0.6359)	(0.6174)	(0.6985)	(0.6731)
FDIRI		−8.3104**		−3.9564	0.2764
		(3.0903)		(4.4056)	(3.6503)
Avg. Tariff		−0.1508**		−0.0863*	−0.1198***
		(0.0510)		(0.0338)	(0.0364)
GDP			−0.0003		−0.0003*
			(0.0002)		(0.0002)
GDP/capita			0.0005***		0.0003***
			(0.0001)		(0.0001)
No. PTAs				0.2157***	0.1532**
				(0.0475)	(0.0509)
Crisis*GPA	2.1030***	2.6484***	1.7190*	1.2033	1.1455
	(0.5547)	(0.6975)	(0.6869)	(0.8556)	(0.8091)
Country FE	Yes	Yes	Yes	Yes	Yes
Observations	660	538	630	538	538
R^2	0.2440	0.2997	0.3917	0.4129	0.4783
Adjusted R^2	0.1886	0.2418	0.3448	0.3631	0.4317

Note: $^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$. Robust standard errors in parentheses.

2.5 Concluding remarks

In this paper we investigate the association between commitments in trade agreements not to discriminate in the award of public procurement contracts and public sector openness following the 2008 financial crisis. We view our findings as a first attempt to explore the role of trade agreements as devices to discipline public procurement policies. To date, arguments regarding the role of the GPA and PTAs that include procurement provisions have mostly been conceptual and normative in nature. There has been very little empirical research on the relative contribution of the GPA and PTAs as commitment mechanisms or on their interaction. To the best of our knowledge, this paper is the first effort to analyze empirically the possible commitment role of the GPA using a cross-country panel dataset.

Our analysis complements previous empirical work on this subject, which has tended to focus on whether the GPA or PTAs result in greater foreign sourcing. The empirical literature on the effect of the trade agreements finds only weak or no evidence that they are associated with a subsequent increase in the share of government procurement allocated

to foreign firms/products.[19] The data we use in this paper suggest that whatever the market access-increasing effects of the GPA and PTAs with enforceable public procurement provisions (that is, taking as given the level of PIP at a point in time) the GPA may serve to constrain backsliding into protectionism when economic times are bad, and that the GPA and PP-PTAs function as partial substitutes in sustaining public sector openness.

The potential role of trade agreements as a commitment device is particularly important for public procurement. As documented by Shingal (2015), home bias increases in recessions. Evenett and Shingal (2016) note that many governments sought to allocate greater funding after 2008 to domestic firms through the procurement process. The post-crisis period therefore provides further evidence confirming previous research that governments have incentives (confront significant pressures) to increase home bias in recessions. This is reflected in the decrease in openness of procurement observed in non-GPA members after 2008. The fact that trends in PIP for GPA and non-GPA member countries were similar before the crisis (towards increased openness) is important in this regard. The divergence in PIP trends after 2008 suggests that commitments by governments not to discriminate against trading partners may have played a role in constraining a greater procurement protectionism. The finding that this relationship is observed most strongly for agreements that are binding and enforceable supports this interpretation.

That the GPA may have served as a device that helped to prevent back-sliding in the post-crisis period is corroborated both by our main regression results controlling for public procurement provisions in bilateral PTAs and by the placebo regression exercise. While the cross-country panel regression analysis does not permit strict causal interpretation, the placebo regression results provide additional support for the robustness of the results. We do not observe that the GPA has a similar association with (changes in) private sector consumption import shares. These are by definition exogenous to GPA provisions, which only cover the public sector. For the private sector PTAs should matter of course, as PTAs as such should promote trade. Once this general role of PTAs is controlled for, there is no relationship between GPA*Crisis and private consumption import shares. Whether the patterns observed in the data will continue to be observed in the future is an open question. Data spanning a longer period of time after 2008 will provide additional evidence.

[19]Examples include Evenett and Shingal (2006), Shingal (2011; 2015), Rickard and Kono (2013) and Kutlina-Dimitrova and Lakatos (2016).

Table A-1: PTAs with and without public procurement provisions

w/o procurement provisions	w/ procurement provisions
APTA	Australia-Chile
APTA-Accession of China	Australia-China
ASEAN - India	Australia-NZ
ASEAN - Japan	Brunei-Darussalam-Japan
ASEAN FTA	CAFTA-DR
ASEAN- Australia/New Zealand	CEFTA
ASEAN-Korea	Commonwealth of Independent States (CIS)
ASEAN–China	Canada - Costa Rica
Agadir Agreement	Canada-Chile
Andean Community	Canada-Colombia
Armenia-Kazakhstan	Canada-Honduras
Armenia-Turkmenistan	Canada-Panama
Central American Common Market	Canada-Peru
CARICOM	Chile - China
CEMAC	Chile - Costa Rica
Common Economic Zone (UKR, BLR, KAZ, RUS)	Chile-Colombia
COMESA	Chile-El Salvador
Canada-Jordan	Chile-Guatemala
Chile - India	Chile-Honduras
Chile - Vietnam	Chile-Japan
Chile-Malaysia	Chile-Nicaragua
Chile-Mexico	China-Korea
Chile-Vietnam	China-Switzerland
China – New Zealand	Colombia-Northern Triangle
China-Hong Kong	Costa Rica - Singapore
China-Macau	Costa Rica-Colombia
Colombia-Mexico	Costa Rica-Peru
Costa Rica-China	Eurasian Economic Union (EAEU)
East African Community (EAC)	EFTA - Albania
EAC-Burundi/Rwanda	EFTA - Jordan
Eurasian Economic Community	EFTA - Lebanon
EAEU-Armenia	EFTA - SACU
EAEU-Kyrgyzstan	EFTA - Serbia
Economic Cooperation Organization FTA	EFTA - Tunisia
ECOWAS	EFTA -Morocco
EU - Albania	EFTA-Bosnia and Herzegovina
EU - Cote d'Ivoire	EFTA-Central America
EU - Lebanon	EFTA-Chile
EU - San Marino	EFTA-Colombia
EU-Andorra	EFTA-Macedonia
EU-FYR Macedonia	EFTA-Mexico
EU-Faroe islands	EFTA-Palestinian Authority
EU-NZ-Malaysia	EFTA-Peru
EU-Papua New Guinea-Fiji	EFTA-Turkey
EU-Syria	EU - Algeria
El Salvador Honduras	EU - Bosnia and Herzegovina
El Salvador-Cuba	EU - Cameroon
Gulf Cooperation Council	EU - Central America
Global System of Trade Preferences Agreement	EU - Jordan
Georgia - Armenia	EU - Montenegro
Georgia - Kazakhstan	EU - Serbia
Georgia - Turkmenistan	EU -Tunisia
Georgia-Azerbaijan	EU- Egypt
Georgia-Russia	EU- Georgia
Georgia-Ukraine	EU- Morocco

Guatemala- Chinese Taipei
India - Afghanistan
India – Bhutan
India – Malaysia
India – Nepal
India – Singapore
India - Sri Lanka
Japan - Indonesia
Japan- Malaysia
Korea-India
Korea-Turkey
Korea-Vietnam
Kyrgyz Republic - Armenia
Kyrgyz Republic Uzbekistan
Kyrgyz Republic-Kazakhstan
Kyrgyz Republic-Ukraine
Kyrgyz republic-Moldova
LAIA-ALADI
Lao-Thailand
MERCOSUR
MERCOSUR - India
Malaysia-Australia
Mauritius-Pakistan
Mexico-Uruguay
Mexico-Panama
Mexico-Peru
Nicaragua-Chinese Taipei
Pan-Arab FTA
Papua New Guinea-Australia
Pakistan - Malaysia
Pakistan - Sri Lanka
Panama - Chinese Taipei
Panama-Chile
Panama-DR
Panama-Nicaragua
Peru - China
Peru-Chile
Russia-Azerbaijan
Russia-Belarus-KZ
Russia-Serbia
Russia-Tajikistan
Russia-Turkmenistan
Russia-Uzbekistan
SACU
SADC
SADC-Seychelles
SAFTA
SAFTA-Afghanistan
SAARC Preferential Trading Arrangement
South Pacific Regional Trade & Econ. Coop. Agreement
Thailand-NZ
Turkey - Albania
Turkey - Chile
Turkey-Mauritius
Ukraine - Azerbaijan
Ukraine - Belarus
Ukraine - Kazakhstan
Ukraine - Tajikistan
EU-CARIFORUM
EU-Chile
EU-Colombia and Peru
EU-Eastern and Southern Africa
EU-Mexico
EU-Palestinian authority
EU-South Africa
EU-Turkey
Egypt – EFTA
Egypt – Turkey
Faroe Island - Norway
Faroe Islands-Switzerland
GCC-Singapore
HK China-Chile
Iceland - Faroe Islands
Iceland-China
India- Japan
Israeli-Mexico
Japan - Australia
Japan - Philippines
Japan - Thailand
Japan - Vietnam
Japan-Mexico
Japan-Mongolia
Japan-Peru
Jordan – Singapore
Korea-Australia
Korea-Chile
Korea-Colombia
Korea-Singapore
Melanesian Spearhead Group
Mexico-Central America
NAFTA
Pacific Island Countries Trade Agreement
Pacific Alliance
Pakistan - China
Panama-Central America
Panama-El Salvador
Panama-Guatemala
Panama-Honduras
Panama-Peru
Panama-SGP
Peru-Korea
Peru-SGP
SGP-Australia
TPP
Thailand - Australia
Trans-Pacific Strategic Economic Partnership
Turkey – Bosnia/Herzegovina
Turkey - Jordan
Turkey - Montenegro
Turkey - Morocco
Turkey - Palestine
Turkey - Serbia
Turkey - Syria
Turkey Israel
Turkey- Georgia
Turkey-Macedonia

Ukraine - Turkmenistan	Turkey-Tunisia
Ukraine - Uzbekistan	US- Jordan
	US-Australia
	US-Bahrain
	US-CAFTA-DR
	US-Chile
	US-Colombia
	US-Morocco
	US-Oman
	US-Panama
	US-Peru
	Ukraine - FYR Macedonia
	Ukraine - Moldova
	WAEMU

Table A-2: Subset of PTAs including procurement language/provisions

w/o enforceable provisions	w/ enforceable provisions
Australia-China	Australia-Chile
Australia-NZ	Canada-Chile
Brunei-Japan	Canada-Colombia
CAFTA-DR	Canada-Honduras
Central European Free Trade Agreement	Canada-Panama
CIS	Canada-Peru
Canada - Costa Rica	Chile - Costa Rica
Chile - China	Chile-El Salvador
Chile-Colombia	Chile-Guatemala
China-Korea	Chile-Honduras
China-Switzerland	Chile-Japan
Colombia-Northern Triangle	Chile-Nicaragua
Costa Rica-Colombia	Costa Rica - Singapore
Costa Rica-Peru	EAEU
EFTA - Albania	EFTA-Central America
EFTA - Jordan	EFTA-Chile
EFTA - Lebanon	EFTA-Colombia
EFTA - SACU	EFTA-Mexico
EFTA - Serbia	EFTA-Peru
EFTA - Tunisia	EU - Central America
EFTA -Morocco	EU- Georgia
EFTA-Bosnia and Herzegovina	EU-CARIFORUM
EFTA-Macedonia	EU-Chile
EFTA-Palestinian Authority	EU-Colombia and Peru
EFTA-Turkey	Faroe Island - Norway
EU - Algeria	Faroe Islands-Switzerland
EU - Bosnia and Herzegovina	GCC-Singapore
EU - Cameroon	HK China-Chile
EU - Jordan	Israeli-Mexico
EU - Montenegro	Japan - Australia
EU - Serbia	Japan-Mexico
EU -Tunisia	Japan-Peru
EU- Egypt	Korea-Australia
EU- Morocco	Korea-Chile
EU-Eastern & Southern Africa	Korea-Colombia
EU-Mexico	NAFTA
EU-Palestinian authority	Panama-SGP
EU-South Africa	Peru-Korea
EU-Turkey	Peru-SGP
Egypt - EFTA	Trans Pacific Economic Partnership
Egypt - Turkey	US-Australia
Iceland - Faroe Islands	US-Bahrain
Iceland-China	US-Chile
India- Japan	US-Colombia
Japan - Philippines	US-Oman
Japan - Thailand	US-Panama
Japan - Vietnam	US-Peru
Japan-Mongolia	Australia-Chile
Jordan - Singapore	Canada-Chile
Korea-Singapore	Canada-Colombia
Melanesian Spearhead Group	Canada-Honduras
Mexico-Central America	Canada-Panama
Pacific Island Countries Trade Agreement	Canada-Peru
Pacific Alliance	Chile - Costa Rica
Pakistan - China	Chile-El Salvador

Panama-Central America	Chile-Guatemala
Panama-El Salvador	Chile-Honduras
Panama-Guatemala	Chile-Japan
Panama-Honduras	Chile-Nicaragua
Panama-Peru	Costa Rica - Singapore
Singapore-Australia	Eurasian Economic Union
Thailand - Australia	EFTA-Central America
Turkey – Bosnia/Herzegovina	EFTA-Chile
Turkey - Jordan	EFTA-Colombia
Turkey - Montenegro	EFTA-Mexico
Turkey - Morocco	EFTA-Peru
Turkey - Palestine	EU - Central America
Turkey - Serbia	EU- Georgia
Turkey - Syria	EU-CARIFORUM
Turkey Israel	EU-Chile
Turkey- Georgia	EU-Colombia and Peru
Turkey-Macedonia	Faroe Island - Norway
Turkey-Tunisia	Faroe Islands-Switzerland
US- Jordan	GCC-Singapore
US-CAFTA-DR	HK China-Chile
US-Morocco	Israeli-Mexico
Ukraine - FYR Macedonia	Japan - Australia
Ukraine - Moldova	Japan-Mexico
WAEMU	Japan-Peru
	Korea-Australia
	Korea-Chile
	Korea-Colombia
	NAFTA
	Panama-SGP
	Peru-Korea
	Peru-SGP
	CPTPP
	Trans Pacific Economic Partnership
	US-Australia
	US-Bahrain
	US-Chile
	US-Colombia
	US-Oman
	US-Panama
	US-Peru

Table A-3: Mundlak correction regression results

	(1)	(2)	(3)	(4)	(5)
GPA	2.9491	0.2323	4.0039**	0.1484	1.5034
	(1.6268)	(1.7397)	(1.5296)	(1.6401)	(1.7796)
Crisis	-0.5585	-0.6908	-1.1117*	-0.7842	-0.8964
	(0.3806)	(0.5236)	(0.4584)	(0.5431)	(0.5456)
GPA*Crisis	1.6694***	1.8130***	1.8708***	1.3929**	1.5017**
	(0.4522)	(0.5068)	(0.5011)	(0.5391)	(0.5380)
FDIRI		2.9076		4.1790	6.1776
		(4.1334)		(4.5108)	(4.2213)
Tariff		-0.0868		-0.0680	-0.0745
		(0.0482)		(0.0492)	(0.0491)
GDP			0.0001		0.0002
			(0.0001)		(0.0001)
GDP/Capita			0.0001***		0.0001*
			(0.0000)		(0.0000)
No. PTAs				0.0628**	0.0389
				(0.0196)	(0.0211)
Avg FDIRI		-7.9432		-4.9917	-4.2315
		(9.4620)		(8.3390)	(9.6724)
Avg Tariff		-0.4946*		-0.4664*	-0.5451**
		(0.2230)		(0.1942)	(0.1889)
Avg GDP			-0.0010***		-0.0010***
			(0.0003)		(0.0003)
Avg GDP/Capita			-0.0002**		-0.0001*
			(0.0001)		(0.0001)
Avg No. PTAs				0.1952	-0.0062
				(0.2068)	(0.1644)
Constant	9.5197***	13.4687***	10.2262***	10.1773***	13.2172***
	(1.4710)	(2.0449)	(1.3685)	(3.0227)	(2.5097)
Observations	660	538	630	538	538

Note: $^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$. Standard errors in parentheses.

Chapter 3

Trade Volume Effects of Public Procurement Provisions in Preferential Trade Agreements

[1]

Abstract

This paper presents novel empirical evidence that the level of detail and legal enforceability of public procurement provisions in preferential trade agreements are associated with a significantly higher volume of bilateral trade flows absorbed by public sector final demand in the importing country. This result is primarily driven by developments in the European Union. The existing literature on the relationship between international public procurement commitments and public sector openness emphasizes the importance on the enforceability of international provisions. Arguably, enforcement of EU public procurement regulation is particulary pronounced. The findings presented in this paper thus complement and confirm the existing literature on the topic.

3.1 Introduction

Public procurement is a large market. In most economies, it makes up more than 10% of GDP (Worldbank 2017). Public administrations thus act not only as market regulators but are themselves important economic actors. Incentives in the government sector however are often different from those of private economic actors. One example is the strong tendency of the state to prefer domestic firms over foreign firms when awarding public procurement

[1] I would like to thank Bernard Hoekman for numerous comments and suggestions.

contracts.[2] This preference materializes in import shares in the public procurement (PP) sector that are substantially lower than in the general economy.

For a long time, public procurement has been sidelined in international trade negotiations. The first version of the GATT Government Procurement Agreement (GPA) enter into force only 1981 and the number of PP provisions in preferential trade agreements (PTAs) started to increase only recently (see Figure 3.1 below and Dengler and Hoekman (2018) for further details). Reducing international discrimination in public procurement is the stated aim of these provisions. Crucially though, the content of PP provisions varies widely across different PTAs, in terms of scope and content as well as in terms of legal enforceability (Shingal and Ereshchenko 2018). Among the most recent PTAs, more and more include enforceable PP commitments with extensive coverage, not confined to reciprocity of market access but also including details on regulatory questions and the practical implementation of the provisions (B. M. Hoekman 2017).[3]

The emergence and proliferation of detailed PP provisions in PTAs raises the question of whether these more substantive types of agreements have a real effects on bilateral trade in a systematic way. In this paper I study to what extent the proliferation and deepening of PP provisions in recent PTAs leads to tangible changes in the tradeflows they cover, using a bilateral gravity framework. Previous research has not found real effects of PP openness commitments and attributed this to a lack of legal enforceability (Rickard and Kono 2014). In this paper I revisit this question using more direct measures of PP imports and a novel classification of PTA PP provisions by Shingal and Ereshchenko (2018), explicitly designed to distinguish PP provisions by their legal enforceability.

I find that the presence of deep PP provisions is indeed associated with significantly higher trade volume absorbed by the government sector in the importing country. When looking at government imports disaggregated by economic sector, it becomes clear that this aggregate finding is driven primarily by service sector imports. Furthermore, the real impact of PP provisions seems to be driven primarily by the European Union. The results are not robust to excluding intra-EU bilateral flows.

There is a considerable literature on estimating the impact of trade policy on bilateral trade flows using gravity-type specifications (see Head and Mayer (2014) for a recent survey).

[2] A particularly prominent example of this are the *Buy American* provisions in the United States. For more systematic evidence, see Breton and Salmon (1996), Evenett and Hoekman (2005) and Shingal (2015).

[3] B. M. Hoekman (2017) gives the example of the *Agreement between New Zealand and Singapore on a Closer Economic Partnership*, which aims to establish a "single New Zealand/Singapore government procurement market" which is going beyond reciprocal access to e.g. establish mutual consistency of national PP tender rules and specifications.

These type of regressions are well suited to investigate potential effects of trade agreements on bilateral trade flows, since for each bilateral flow it can be directly controlled if and by which type of agreement it is governed. In this spirit, Dhingra, Freeman, and Mavroeidi (2018) investigate the effect of deep vs shallow EIAs, exploiting recent information from the DESTA database on the presence of different types of provisions in trade agreements. Their paper focuses on the trade policy implications of Brexit for the UK. The underlying trade flow data originates from the World Input Output Database (WIOD). This enables the authors to investigate both gross versus value added trade flows and to split flows by intermediate and final use in the importing country. However, Dhingra, Freeman, and Mavroeidi (2018) do not attempt to split final demand further into private and government final demand and do not investigate PTA provisions governing the government sector at all.

Dengler and Hoekman (2018) investigate the effect of the GPA and PTA PPs on public sector openness and find that PTA PPs, in particular their enforceability, have a significant effect on average public consumption import shares following the 2008 financial crisis. Dengler and Hoekman (2018) employ a country panel methodology, which does not directly link bilateral trade flows to the potential presence of an agreement. The results are therefore suggestive but raise the question of whether the measured effects are robust to different methodologies.

This paper investigates the effect of public procurement provisions using bilateral gravity regressions, exploiting the novel PTA public procurement clause classification by Shingal and Ereshchenko (2018). Like Dhingra, Freeman, and Mavroeidi (2018), trade flow data is adopted from WIOD. WIOD enables the calculation of value added trade flows, which is particularly relevant as the indirect imports into the government sector are much higher than direct imports (Ramboll and Chur (2011), Cernat and Kutlina-Dimitrova (2016)). While country-industry import flows absorbed by government consumption demand are partially imputed in WIOD, flows are estimated using an internationally consistent methodology. This is crucial for cross-country comparisons such as this paper. The requirement of international consistency is ruling out the use of micro data on procurement, which exist only for few countries and years. The trade effects of PTAs have been studied previously by Rickard and Kono (2014). They rely on changes in the elasticity of imports to procurement spending as an indirect measure of the trade effects. The availability of direct flows, if partially imputed, through international input output tables makes direct estimation of the trade effects of PP provisions possible.

Rickard and Kono (2014) find no measureable effects of PTA PP provisions on the import elasticity of trade. They explain this with a lack of legal enforceability of the PP provisions and the erection of non-tariff barriers in public procurement in response

to reductions in direct tariff barriers. A novel classification of PP provisions in PTAs enables me to test this conjecture directly. Shingal and Ereshchenko (2018) analyze the public procurement provisions in all PTAs notified to the WTO up to March 2017, using a detailed questionnaire. They then use their detailed qualitative assessment to classify agreements into three broad categories according to the 'depth' or legal enforcability of these provisions.[4] This classification opens up the possibility of systematically investigating the trade effects of those agreements quantitatively.

I estimate the effect of PP provisions as classified by Shingal and Ereshchenko (2018) on value added trade flows absorbed by government consumption constructed from WIOD, using the Poisson Pseudo Maximum Likelihood (PPML) estimator proposed by Santos Silva and Tenreyro (2006). I include importer-time, exporter-time and importer-exporter pair fixed effects in the specification. These account succinctly and comprehensively for a range of standard controls in gravity estimation (Y. V. Yotov et al. 2016). In the main specification, bilateral links covered by enforceable agreements are associated with statistically significantly higher trade volume, relative to links covered by shallow provisions or no PTA at all. Investigating the effects on sectoral levels, it becomes clear that aggregate flows are driven primarily by service sector imports.

A large share of bilateral links in the baseline sample are trade flows between member states of the European Union (EU), due to the WIOD country coverage. The EU negotiates all trade agreements on behalf of its member countries at the supra-national level. In addition, procurement rules in the EU are designed to foster non-discrimination and are viewed as part of the EU's Single Market (Cantore and Togan 2017). This sets the EU procurement rules appart from the PP provisions of other PTAs. I therefore reestimate the baseline results on an additional, smaller sample that treats inner-EU trade flows as domestic and focuses on extra-EU trade flows between the EU and third countries only. The statistical significance of the association between enforceable PP provisions and the size of the underlying trade flow vanishes completely. Inspecting the composition of bilateral flows over time, it becomes apparent that the variation in the baseline sample is primarily driven by intra-EU flows.[5]

These results are consistent with prior findings in the literature. Rickard and Kono (2014)

[4]In the following, I will use the terms deep provisions and (legally) enforceable provisions interchangeably. In the detailed questionnaire, Shingal and Ereshchenko (2018) explicitly cover legal enforceability. More than 90% of the agreements included in their deep PP provisions group contain provisions to ensure legal enforceability.

[5]In particular, variation is driven by the accession of new member states to the EU throughout the sample period, especially in 2004 (Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia and Slovenia) and 2007 (Bulgaria and Romania).

emphasize the level of enforceability as crucial. Enforceability within the EU is higher than in other international agreements, as with the European Commission there is a joint agency ensuring compliance of member states and an EU-level judicial system exists. Tas et al. (2017) study procurement outcomes in the European Economic Area (EEA), Switzerland and Macedonia for tenders covered by GPA rules using micro-level tender data released by the EU and find that procurement is more open than for tenders not covered by the GPA. Cantore and Togan (2017) analyze EU procurement regulation and view it as generally consistent with GPA rules. Three out of four PTA PP provisions classified as deep Shingal and Ereshchenko (2018) are categorized as en par or exceeding the level of coverage prescribed by the GPA. I view this as additional evidence that legal enforceability of provisions is indeed crucial for seeing real effects on trade flows.

The body of this paper is organized as follows. In Section 3.2, I present more information on the Shingal and Ereshchenko (2018) PTA classification and the trade flow data from the World Input Output Database. The empirical methodology and baseline results are presented in Section 3.3. Section 3.4 presents the results on the second sample, considering only EU tradeflows with third parties. Section 3.5 concludes. Two appendices contain additional detail on WIOD and present several additional robustness checks.

3.2 Data

3.2.1 Shingal and Ereshchenko (2018) PTA Classification

The main goal of this paper is to assess the importance of enforceability of procurement provisions on the actual trade impact of the agreement. To this end, I make use of a novel classification of public procurement provisions present in PTAs by Shingal and Ereshchenko (2018). Their classification is based on a questionnaire approach including 100 specific questions across 8 broad categories. Responses are either binary or detailed, enabling a classification of PP clauses at both extensive and intensive margins.[6] Based on the detailed questionnaire, Shingal and Ereshchenko (2018) classify the agreements into three groups, enabling quantitative comparison of the relative coverage and depth of the PP provisions across PTAs. The novelty of their approach comes from combining a in-depth qualitative assessment of PP clauses, similar to Anderson et al. (2011) and Ueno (2013), with a classification amenable to quantitative use. Dür, Baccini, and Elsig (2014) also categorize PP provisions in an index in their DESTA database, however, their assessment is based on a much smaller number of questions.

[6]Their 8 categories, with the respective number of questions in parantheses, are *Overview* (4), *Non-discrimination* (14), *Coverage* (40), *Ex-ante transparency* (3), *Procedural disciplines* (26), *Ex-post transparency* (4), *Dispute settlement* (4) and *"New" issues* (4).

Table 3.1: PTA Classification Sample

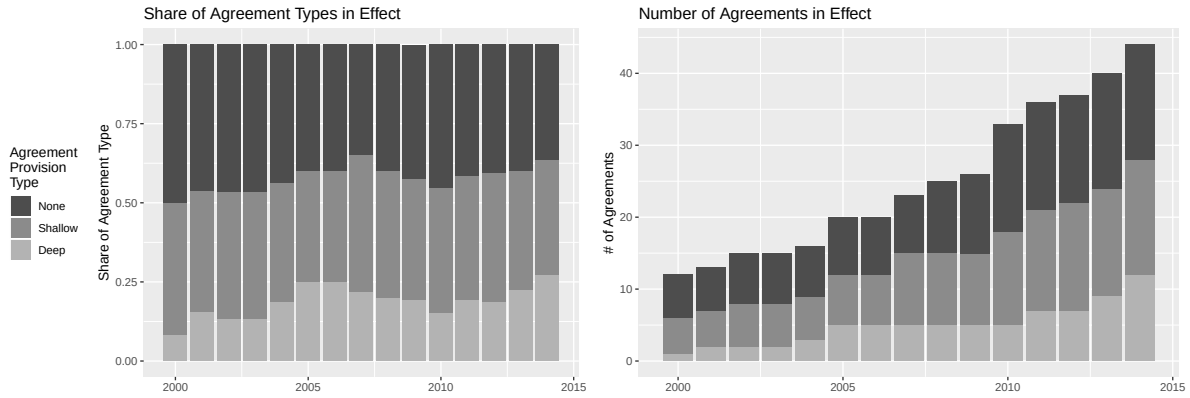
provisions	enforceable	count	share
No	No	16	0.36
Yes	No	17	0.38
Yes	Yes	12	0.27
Total		45	1.01

Note: Shingal and Ereshchenko (2018) agreements involving at least two countries covered by WIOD, 2016 revision. Deviations of total share from 1 due to rounding errors.

Shingal and Ereshchenko (2018) provide a broad classification of PTAs by the type of PP provisions they include. PTAs are classified into three groups. First, PTAs without reference to government procurement and containing no PP provisions. Second, PTAs which include either brief references to PP or a full chapter, but without legally binding text (shallow PP provisions). Third, agreements including detailed PP provisions or explicitly referencing the inclusion of GPA provisions (deep PP provisions).

A disaggregation by PP classification of the number of agreements in the sample is presented in Table 3.1. In principle, the Shingal and Ereshchenko (2018) database covers all PTAs which have been notified to the WTO up until March 2017, however the associated trade flow data used in this paper is based on data from the WIOD, 2016 revision. In this paper, I will therefore focus on the subset of agreements covered by Shingal and Ereshchenko (2018) which entered into force up until December 2014 and involve at least two countries in the WIOD sample, i.e. only agreements which cover some of the bilateral flows in the sample. This criterion results in a sample of 45 agreements. The agreements covered are about evenly distributed by PP classification, with agreements featuring deep or enforceable PP provisions having the lowest share at about 27%.

Figure 3.1: Public Procurement Provisions Over Time



While Table 3.1 presents shares over the entire WIOD sample period from 2000 to 2014, Figure 3.1 plots the number and relative shares of agreement types which are currently in force for all years of the sample separately. The absolute number of PTAs in force expands throughout the sample period both overall and for all classifications. The relative share of agreements including shallow or deep PP provisions increases throughout the sample period, however, not uniformely and not substantially.

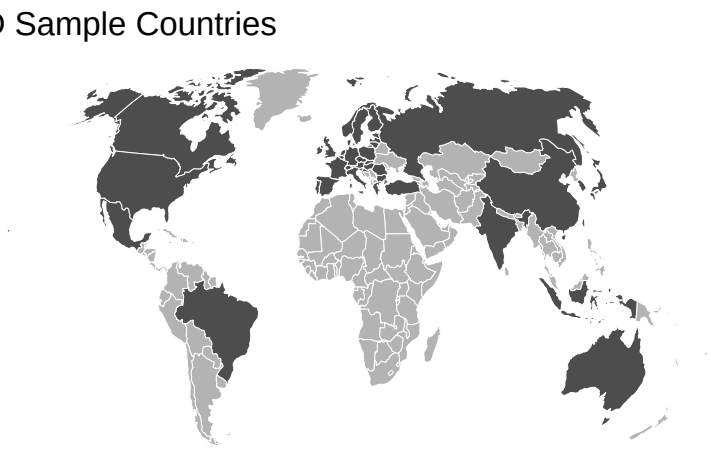
The patterns in Figure 3.1 imply that while the absolute number of PTAs with deep provisions increases, their share among agreements currently in force does not raise substantially. Importantly, here all agreements are weighted equally, irrespective of the number of member countries or the size of trade flows governed by the agreement. When disaggregating bilateral trade flows by type of agreement covering them, the picture changes (see Subsection 3.2.3 below).

3.2.2 World Input Output Database

The World Input Output Database (M. P. Timmer et al. 2015) combines national accounts data and data on trade flows to estimate full international input output tables, including international industry-to-industry flows and a split of final demand categories by country-industry source. The WIOD, revision 2016 data covers 43 countries, plotted in Figure 3.2.[7] Despite the limited country coverage, WIOD covers more than 85% of World GDP throughout the sample period (Dengler and Hoekman 2018).

[7]The differences of the 2016 revision of WIOD relative to the first release are described in further detail by M. Timmer et al. (2016).

Figure 3.2: World Input Output Database Country Coverage



In this paper, the split of final demand categories by country-industry source is of particular importance, since it enables an analysis of public sector consumption demand disaggregated by country. WIOD data has a couple of advantages over other datasets. Since WIOD features complete international input output flows, it is possible to calculate value added imports by final demand category in addition to direct gross imports. Other databases, e.g. GTAP, also split final demand categories by country source, however, due to the lack of complete international input output tables, it is not possible to compute value added imports. For other databases that do include value added trade, e.g. OECD TiVA, coverage extends only up to 2011, whereas the latest release of WIOD covers flows up to 2014.

The main analysis is done on value added import flows rather than gross import flows. Gross imports from Country A to Country B are simply direct bilateral flows from A to B absorbed by public final consumption demand in B, aggregated over all industries. Value added imports are tracking the actual foreign content embedded in public final consumption demand both foreign and domestic. One general advantage of using value added imports is avoiding double-counting of trade flows in the case of re-exports.

In the context of public procurement, value added flows have a second advantage. Relying on gross import flows only captures *direct* imports of the government sector. However, in practice indirect imports, e.g. via domestic subsidiaries of foreign companies are magnitudes larger than direct imports.[8] The value added flows do not directly correspond to foreign ownership measures. However, to the extend to which foreign subsidiaries are more likely

[8]Ramboll and Chur (2011) report for the EU that on average in 2009 just 1.4 percent of tenders were awarded to foreign bidders directly, while 11.4 percent were awarded indirectly. In terms of tender value, the numbers are equally striking with 3.4 percent directly and 15.7 percent indirectly awarded to foreign bidders.

Table 3.2: Government Consumption Industry Decomposition

	Total Flow Decomp.		Import Flow Decomp.		Foreign Content Shares	
industry	GO	VA	GO	VA	GO	VA
man	1.27	7.10	58.26	37.84	53.97	33.69
ser	98.08	88.22	40.87	47.70	0.49	3.42
oth	0.65	4.68	0.86	14.46	1.57	19.53
tot	100.00	100.00	100.00	100.00	1.17	6.32

Data: WIOD, revision 2016. Note: The 'Total Flow Decomposition' columns show the decomposition of total government consumption by economic sector, the 'Import Flow Decomposition' columns show the decomposition of government consumption imports. The 'Foreign Content Shares' columns show the share of imports in the total government consumption demand from the respective industry. All shares are GDP-weighted sample averages. The service sector is defined as the aggregate of ISIC Rev. 4 industries G - S.

to source from their parent companies abroad than domestic bidders, this will be reflected in the value added import flows. More directly, PP clauses preventing application of domestic rules of origin type regulations will be directly reflected in value added import flows (of both domestic companies and foreign subsidiaries), while these type of changes will not be captured at all in gross import flows.

The calculation of value added imports follows the methodology of Koopman, Wang, and Wei (2014). The authors propose a new method for decomposing gross trade flows into several value added compontents. For this paper, the foreign value added component is of interest. Further details on the construction of variables and some additional stylized facts are presented in Appendix 3.A.

Table 3.2 shows total government consumption and imports disaggregated by industry plus the general foreign content shares of the different sectors for both gross output and value added. The first two columns show the strong orientation of government consumption expenditures towards services. Over 98 percent of overall gross expenditures and over 88 percent of value added are spent on services. In contrast, when looking at import flows only, manufactured goods make for a much larger share of the overall share. In the case of gross output, the share is even larger than that of services.

Table 3.3: Bilateral Trade Flow Sample

flow type	agreement	provisions	enforceable	count	share
domestic	-	-	-	645	0.02
international	No	No	No	16642	0.60
international	Yes	No	No	686	0.02
international	Yes	Yes	No	1464	0.05
international	Yes	Yes	Yes	8298	0.30
Total				27735	0.99

Note: Bilateral trade flows in WIOD sample, pooled over all sample years (2000 – 2014) and disaggregated by Shingal and Ereshchenko (2018) agreement classification coverage. Deviation of total share from 1 due to rounding errors.

3.2.3 Bilateral trade flows

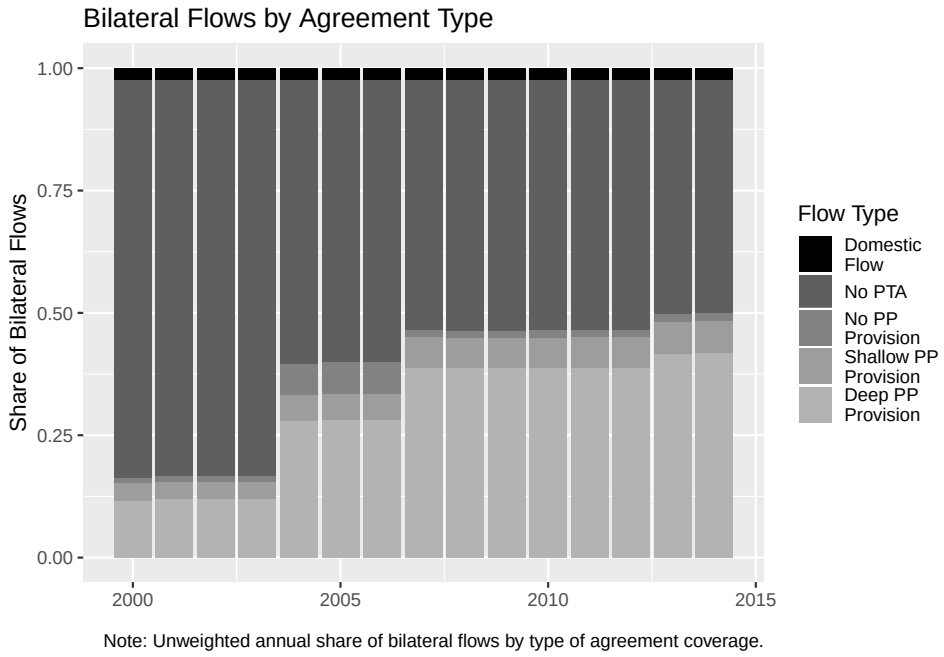
For the regressions in this paper, information on bilateral public consumption imports from WIOD are complemented by information on the kind of agreement covering the respective bilateral flow from the Shingal and Ereshchenko (2018) classification. A disaggregation of the bilateral flows by the type of agreement covering the respective flow is presented in Table 3.3.

While the relative number of agreements by type in the sample is relatively balanced (see Table 3.1), the relative number of bilateral trade flows is biased towards flows which are not governed by any agreement and flows governed by agreements with deep and enforceable PP provisions. This difference stems from two characteristics of the data. First, agreements differ in the number of member countries. Agreements with more members will translate into an over-proportional number of bilateral flows. Second, the WIOD data is biased towards European countries, since it covers all EU members. Since trade relations are identical for all EU member countries, agreements involving the EU are 'replicated' in the bilateral flows data for each EU member country. In order to address the second feature, the main results are replicated on a bilateral flow sample treating the EU as a single country in Section 3.4.

Figure 3.3 decomposes this data further for each year of the sample. It can be seen that the share of bilateral trade flows governed by deep and enforceable agreements is increasing between 2000 to 2014 from 12% to over 42% of all international flows. This stands in contrast to the stagnant share of the overall relative number of agreements (see Figure

3.1), implying that deep agreements are becoming larger over time.

Figure 3.3: Agreement Coverage of Bilateral Flows Over Time



3.3 Estimation

In this Section I present the bilateral gravity regression exercise. The estimation strategy will be outlined in the next Subsection, before results are presented subsequently.

3.3.1 Estimation Strategy

The aim of the analysis is to test to which extent deep and legally enforceable PP provisions have an impact on openness of government consumption demand. The gravity model has emerged as a workhorse model for this type of questions (Y. V. Yotov et al. (2016)). The bilateral gravity framework is consistent with the main theoretical models of international trade and directly maps theoretical counterparts into the available data.

For the baseline specification, I estimate the effect of PTA PP provisions on bilateral trade flows using the Santos Silva and Tenreyro (2006) PPML estimator, as recommended by Y. V. Yotov et al. (2016). This differs from Dhingra, Freeman, and Mavroeidi (2018), who employ a fixed-effects (FE) estimator instead.

The PPML estimator has several advantages over FE. First, it enables the researcher to include zero flows in the estimation, which would have to be dropped in the log-linearized FE specification. In public sector imports, in particular gross imports, zero flows are more prevalent than in general trade data, making this concern particularly salient. Second, the

PPML estimator takes care of potential heteroscedasticity of trade flows. Heteroscedasticity is problematic for the FE estimator, since the log-transformation does not treat positive and negative errors symetrically. Third, Larch et al. (2017) provide a new, computationally efficient implementation of the PPML estimator, reducing the computational burden in the presence of high-dimensional fixed effects.

Based on Larch et al. (2017), the reduced form equation to be estimated is given by

$$X_{ijt} = \exp\left(\phi_{it} + \psi_{jt} + \gamma_{ij} + \boldsymbol{\beta TA_{ijt}}\right) + \varepsilon_{jit} \tag{3.1}$$

where X_{ijt} is the (gross or value added) part of the overall trade flow from Country i to Country j that is absorbed by government consumption demand in Country j at time t, ϕ_{it} is an exporter-time fixed effect, ψ_{jt} is an importer-time fixed effect, γ_{ij} is a country-pair fixed effect and $\boldsymbol{TA_{ijt}}$ is a vector of policy variables describing the trade regime governing the respective bilateral flow. The overall effect of the agreement $\boldsymbol{TA_{ijt}}$ is disaggregated into three indicator variables, indicating the presence of a gerneral trade agreement (independently from PP coverage), presence of a public procurement provisions (shallow or deep) in the agreement and presence of deep and enforceable public procurement provisions as follows

$$\boldsymbol{\beta TA_{ijt}} = \beta_1 Agreement_{jit} + \beta_2 Provisions_{jit} + \beta_3 Enforceable_{jit}. \tag{3.2}$$

The variables are thus defined as subsets of their respective predecessors, i.e. $Enforceable$ is always zero if $Provisions$ is zero and $Provisions$ is always zero if $Agreement$ is zero. This approach follows the setup of Dhingra, Freeman, and Mavroeidi (2018).

The effects of PTA provisions on bilateral trade flows are likely to materialize fully at medium to low frequency, since agents need time to adjust to the change of rules. I therefore follow the convention in the bilateral gravity literature and estimate the baseline specification at 4 year intervals. In addition, this isolates the estimates from high frequency annual variation. In order to exploit the recency of the WIOD 2016 revision, I include data for 2014 in every specification. The baseline sample thus includes bilateral trade flows for 2002, 2006, 2010 and 2014.[9]

[9]For comparison to Dhingra, Freeman, and Mavroeidi (2018)'s baseline specification of two year intervals, results from reestimating the baseline specification at this higher frequency are presented in Appendix **??**. While the coefficients are quantitatively smaller, the main qualitative patterns continue to hold. This incidentially lends support to the adjustment time hypothesis motivating the use of lower frequency data by the literature in the first place.

Including a full set of country-time and country-pair fixed effects precludes the inclusion of country membership in the WTO GPA in the policy vector in equation **??**. Membership in the GPA has been virtually constant over the sample period, implying that GPA membership is not identified in the presence of country-pair fixed effects.[10]

3.3.2 Baseline Results

The main estimates of interest are the policy parameters on the different types of agreement. As these are indicator variables, the trade volume effect in percentage terms is calculated from the estimates as

$$(e^{\hat{\beta}_{policy}} - 1) \cdot 100 \tag{3.3}$$

where $\hat{\beta}_{policy}$ is the respective policy estimate (Y. V. Yotov et al. (2016)). The joint effect of two types of policies can be obtained by summing the point estimates as

$$(e^{\hat{\beta}_{p1}+\hat{\beta}_{p2}} - 1) \cdot 100. \tag{3.4}$$

Baseline regression results for total value added import flows into the government sector are presented in Table 3.4. Regressing bilateral trade flows on the PTA classficiations separately in Models (1) to (3) yields the following results. When calculating the estimated percentage effect using equation **??** for all PTAs – PP provisions or not – with 25.83 percent the effect is smaller than when looking only at those PTAs with shallow provisions, with 29.7 percent, or those with deep provisions, which are associated with a 31.33 percent trade volume increase over trade flows not covered by any agreement. This is consistent with the interpretation that the type of provision actually matters.

The overall effect of PP provisions in specification (4) is large and jointly significant. Relative to flows not covered by a PP provision, government consuption import flows are higher by 31.33 percent, using the transformation **??**. When looking at the disaggregation between shallow and deep provisions, it can be seen that the coefficient on enforceable PP provisions is much larger than the coefficient on shallow provisions and statistically significant, albeight only at the 5% level. While shallow provisions are not associated with trade volume significantly higher than trade flows not governed by any form of PP provisions, deep provisions are estimated to go along with a 20.94 higher import volume

[10]The sole exceptions are the EU accession countries. GPA accession is coincidental with EU accession. Given the large changes in PP rules and regulations associated with becoming an EU member, potential specific effects of GPA membership are arguably impossible to identify separately.

Table 3.4: Total imports - VA - 4 year frequency

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2298***				0.0120
	(0.0391)				(0.1007)
provisions		0.2600***		0.0932	0.0814
		(0.0253)		(0.0898)	(0.1350)
enforcable			0.2726***	0.1902*	0.1970*
			(0.0262)	(0.0933)	(0.0942)
N	7396	7396	7396	7396	7396
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

by the importer's government consumption. When taking into account in addition any type of PTA in Specification (5), the size of the effect rises slightly to 21.77 percent while significance levels remain the same.

Tables 3.5 and 3.6 present the baseline results for services and manufacturing flows separately. The aggregate results are clearly driven by the service sector imports. This is to be expected, as Table 3.2 shows that services trade dominates the overall trade flows into the government sector. While qualitatively the services flows are very similar to the aggregate flows, quantitatively they are larger. For manufacturing flows, enforceability of PP provisions does not appear to be associated with higher trade volumes. According to the Shingal and Ereshchenko (2018) classification, 65% of shallow agreements do not cover services in addition to manufacturing. This might explain why the effects of shallow agreements bear out more strongly when looking at manufacturing trade flows only.

Table 3.5: Services imports - VA - 4 year frequency

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2697***				0.0561
	(0.0322)				(0.0585)
provisions		0.2841***		0.0557	−0.0000
		(0.0298)		(0.1231)	(0.1361)
enforcable			0.2992***	0.2542*	0.2726*
			(0.0306)	(0.1266)	(0.1272)
N	7396	7396	7396	7396	7396
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

Table 3.6: Manufacturing imports - VA - 4 year frequency

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2698***				0.0830
	(0.0568)				(0.1115)
provisions		0.2994***		0.1828**	0.1005
		(0.0459)		(0.0591)	(0.1245)
enforcable			0.3107***	0.1336	0.1474
			(0.0499)	(0.0730)	(0.0753)
N	7396	7396	7396	7396	7396
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

3.4 EU as a single entity

Trade flows are generated at the national level and procurement is a domain of national and sub-national government entities. Thus, the main analysis is focused on investigating the relationship between government consumption import and international agreements on the national bilateral level. Generally, trade agreements are negotiated on the national level as well, making the country-level perspective the natural choice.

An important exception to the national responsibility for PP and trade agreements is the European Union. In the context of this paper, the EU is particularly relevant, since EU member states constitute the majority of the WIOD sample countries with 28 out of 43 countries in total (c.f. Figure 3.2).[11] For EU member countries, decision making potentially differs from the national level for two reasons relevant to this paper. First, all trade agreements are exclusively negotiated at EU level, i.e. individual member states are unable to separately negotiate any form of agreement. Second, the EU has strong non-discrimination clauses in public procurement regarding firms from other member states as part of the EU common market. Assuming the EU PP clauses are perfectly binding, all member states de facto constitute a single PP market. From perspective taken in this paper the EU could be interpreted as a single country, as both the procurement market and trade agreements comprise all EU member countries simultaneously.

In addition to the 'EU as a country' argument just outlined, including the EU member countries separately potentially creates quantitative biases in the baseline results for the following reasons. First, a large part of the agreements covering sample links will involve the EU. This could potentially be of concern, if the EU negotiates all its agreements in a similar way, implying less variation in type of agreements observed in the sample. Second, all agreements involving the EU are 'multiplied' by the number of EU member states. Since level of observations are bilateral links, this gives agreements involving the EU a much higher weight in the overall number of observations, potentially biasing the results.

This section therefore presents the baseline specifications re-estimated on an EU sample. Here, all flows from EU members to non-members have been collapsed and trade flows between EU members are treated as 'domestic' EU flows. In order to explore the concerns of sample bias and lack of variability, the next Subsection presents a set of summary statistics for this sample before regression results are presented in the subsequent Subsection.

Table 3.7: Bilateral Trade Flow Sample

flow type	agreement	provisions	enforceable	count	share
domestic	-	-	-	240	0.06
international	No	No	No	2848	0.74
international	Yes	No	No	404	0.11
international	Yes	Yes	No	132	0.03
international	Yes	Yes	Yes	216	0.06
Total				3840	1.00

Note: Bilateral trade flows in WIOD sample, pooled over all sample years (2000 – 2014) and disaggregated by Shingal and Ereshchenko (2018) agreement classification coverage. Deviation of total share from 1 due to rounding errors.

3.4.1 Summary Statistics on EU level

A disaggregation of bilateral trade flows by agreement coverage type is presented in Table 3.7. The overall number of bilateral flows is 3840, substantially lower than in the baseline sample with over 27,000. The difference reflects the potential size of the bias and stems from the nature of the bilateral flows, the number of which grows exponentially in the number of partners. Compared to the baseline sample (presented in Table 3.3), an even larger share of the observations is not covered by any type of agreement. Among the bilateral flows covered by agreements, the observations are now more evenly distributed among no PP provisions, shallow and deep provisions. This reflects the removal of intra-EU flows, which have been classified as being covered by a PTA with deep provisions.

[11]The WIOD project was originally funded by the European Commission through its 7th Framework Program, explaining the bias towards European countries.

Figure 3.4: Agreement Coverage of Bilateral Flows Over Time

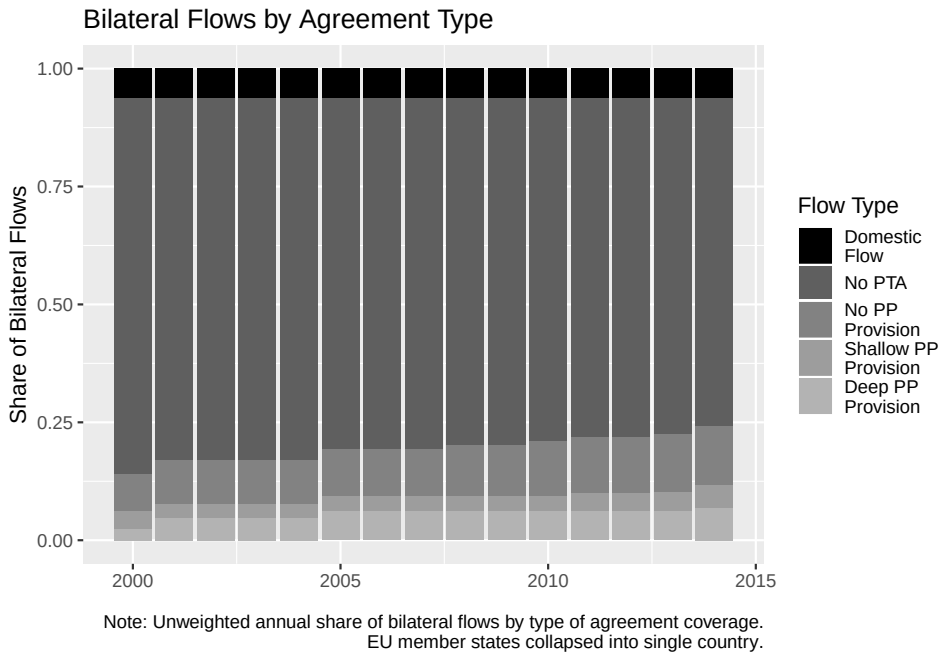


Figure 3.4 further disaggregates the distribution of flows over time. While the share of bilateral observations covered by international agreements with various PP classifications continuously grows, the intertemporal variability is substantially lower than in the baseline sample.

Inspecting the bilateral flow distribution of the EU sample reveals mixed implications. On the one hand, the EU bias in terms of share of observations is removed, at the cost of a much smaller sample. On the other hand, the resulting sample seems even less balanced than the baseline sample, in terms of both cross-sectional and inter-temporal variability between agreement types covering the flows.

3.4.2 Regression results on EU level

Regression results for the EU sample are presented in Table 3.8 for total bilateral flows, Table 3.9 for manufacturing sector import flows and Table 3.10 for service sector import flows.

The results are starkly different from the baseline case – statistical significance basically vanishes from all specifications. In models (4) and (5), the presence of enforceable provisions now has a negative sign, albeight insignificant as well. The results put into question the robustness and the interpretation of the relationships seen in the baseline sample. At the same time, the summary statistics on this robustness sample discussed in the previous Subsection indicate that the results should be interpreted with care.

Table 3.8: results-eu-va-gov-tot-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	−0.0124				−0.1487
	(0.0745)				(0.1266)
provisions		0.0213		0.1240	0.2724
		(0.0507)		(0.1170)	(0.1721)
enforcable			0.0082	−0.1239	−0.1234
			(0.0428)	(0.1239)	(0.1241)
N	1024	1024	1024	1024	1024
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.9976	0.9955

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

In summary, the EU sample robustness exercise qualifies the baseline results. At the same time, also the robustness sample suffers from a lack of variability in the main explanatory variables. Overall, this suggests that the WIOD sample is not idealy suited to study the effect of PTA PP provision classification on public sector openness. In particular, a larger set of countries with appropriate data would be desirable. Since so far, this is not available, the baseline results of this paper should be taken with a grain of salt.

Table 3.9: results-eu-va-gov-man-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.0112				−0.0831
	(0.0790)				(0.1458)
provisions		0.0628		0.1839*	0.2669
		(0.0617)		(0.0810)	(0.1654)
enforcable			0.0178	−0.1836	−0.1831
			(0.0624)	(0.1003)	(0.1004)
N	1024	1024	1024	1024	1024
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.9954	0.9901

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

Table 3.10: results-eu-va-gov-ser-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.0155				−0.0451
	(0.0589)				(0.0979)
provisions		0.0213		0.1140	0.1523
		(0.0695)		(0.1579)	(0.1828)
enforcable			0.0142	−0.1139	−0.1069
			(0.0619)	(0.1682)	(0.1653)
N	1024	1024	1024	1024	1024
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.9987	0.9958

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

3.5 Concluding Remarks

This paper investigates whether bilateral trade links covered by deep and legally enforceable PP provisions are associated with higher trade flows absorbed by government consumption in the importing country using a new classification of PP provisions and value added public sector trade flows constructed from the World Input Output Database. I find that enforceability is indeed associated with significantly higher trade flows, however, this result is driven primarily by developments in the European Union. Once excluding intra-EU trade, the association between PP provision type and size of the underlying trade flow vanishes.

This result is consistent with the existing literature investigating the effects of PP provisions on trade. Using a more indirect measure of government imports, Rickard and Kono (2014) find no effect of PTA PP provisions on trade. They attribute this to non-tariff barriers and a lack of legal enforceability of the PTA provisions. Tas et al. (2017) investigate procurement tenders in the EU and do find tangible effects of the GPA. The results presented in this paper further strengthen the hypothesis that legal enforceability is crucial in achieving tangible results on trade flows from international commitments.

The analysis in this paper is subject to some caveats. First, a robustness exercise in Appendix 3.B.3 using a placebo variable regression does not strengthen a potential causal mechanism for the baseline regression results. Reestimating the baseline regressions on *private* consumption value added imports (arguably unaffected by PTA PP provisions) generates qualitatively very similar results to the baseline estimation. The results from this exercise should be interpreted with care, since the Shingal and Ereshchenko (2018) classification covers only the PP provisions of the respective PTAs. To the extend that PTAs with deeper PP provisions are also those PTAs with deeper provisions regarding the private sector, the results from the placebo exercise could be driven by omitted variable bias.

Second, the World Input Output Database might not be the ideal source of trade flow data, due to its bias towards European countries. In particular South America and Asian countries are not sufficiently covered, limiting the number of PTAs included in the sample. The choice of WIOD data for this analysis is motivated by the necessity of having to split trade flows by the nature of final demand absorbing them in the importing country. Given that other datasets covering a more comprehensive set of countries, such as the GTAP project, do not feature full international input-output tables at the time of writing, WIOD currently provides the most comprehensive possibility for testing the hypothesis of this paper.

Finally, a number of agreements have not been included in the sample as they have been

coming into effect only in the years following 2014, the final sample year of WIOD, 2016 revision. A more complete assessment of the effectiveness of PTA PP provisions will have to await the availability of more up-to-date and geographically comprehensive data in the coming years.

References

Anderson, R, AC Muller, OL Kodjo, J Pardo De Leon, and P Pelletier. 2011. "Government Procurement Provisions in Regional Trade Agreements: A Stepping Stone to Gpa Accession. Chapter 20 in the Wto Regime on Government Procurement: Recent Developments and Challenges Ahead."

Breton, Albert, and Pierre Salmon. 1996. "Are Discriminatory Procurement Policies Motivated by Protectionism?" *Kyklos* 49 (1). Wiley Online Library: 47–68.

Cantore, Carlo Maria, and Sübidey Togan. 2017. "Public Procurement in the Eu." In *The Internationalization of Government Procurement Regulation*, edited by Bernard M. Hoekman Aris Georgopoulos and Petros C. Mavroidis. Oxford: Oxford University Press.

Cernat, Lucian, and Zornitsa Kutlina-Dimitrova. 2016. "TTIP and Public Procurement: Going Beyond the Tip of the Iceberg." CEPS Policy Brief, No. 339.

Dengler, Benedikt, and Bernard Hoekman. 2018. "The Wto Government Procurement Agreement as a Commitment Device: A First Economic Appraisal."

Dhingra, Swati, Rebecca Freeman, and Eleonora Mavroeidi. 2018. "Beyond Tariff Reductions: What Extra Boost from Trade Agreement Provisions."

Dür, Andreas, Leonardo Baccini, and Manfred Elsig. 2014. "The Design of International Trade Agreements: Introducing a New Dataset." *The Review of International Organizations* 9 (3). Springer: 353–75.

Evenett, Simon J, and Bernard M Hoekman. 2005. "Government Procurement: Market Access, Transparency, and Multilateral Trade Rules." *European Journal of Political Economy* 1 (21): 163–83.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics*, 4:131–95. Elsevier.

Hoekman, Bernard M. 2017. "International Cooperation on Public Procurement Regulation." In *The Internationalization of Government Procurement Regulation*, edited by Bernard M. Hoekman Aris Georgopoulos and Petros C. Mavroidis, 568–602. Oxford: Oxford University Press.

Koopman, Robert, Zhi Wang, and Shang-Jin Wei. 2014. "Tracing Value-Added and Double Counting in Gross Exports." *American Economic Review* 104 (2): 459–94.

Larch, Mario, Joschka Wanner, Yoto Yotov, and Thomas Zylkin. 2017. "The Currency

Union Effect: A Ppml Re-Assessment with High-Dimensional Fixed Effects."

Ramboll, Management Consulting, and HTW Chur. 2011. "Cross-Border Procurement Above Eu Threshold." Ramboll, Copenhagen.

Rickard, Stephanie J, and Daniel Y Kono. 2014. "Think Globally, Buy Locally: International Agreements and Government Procurement." *The Review of International Organizations* 9 (3). Springer: 333–52.

Santos Silva, JMC, and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88 (4). MIT Press: 641–58.

Shingal, Anirudh. 2015. "Econometric Analyses of Home Bias in Government Procurement." *Review of International Economics* 23 (1). Wiley Online Library: 188–219.

Shingal, Anirudh, and Viktoriya Ereshchenko. 2018. "Mapping Government Procurement Provisions in Preferential Trade Agreements:evidence from a New Classification."

Tas, Bedri Kamil Onur, Kamala Dawar, Peter Holmes, and Sübidey Togan. 2017. "Does the World Trade Organization Government Procurement Agreement Deliver What It Promises?"

Timmer, Marcel P, Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J De Vries. 2015. "An Illustrated User Guide to the World Input–output Database: The Case of Global Automotive Production." *Review of International Economics* 23 (3). Wiley Online Library: 575–605.

Timmer, Marcel, Bart Los, Robert Stehrer, and Gaaitzen de Vries. 2016. "An Anatomy of the Global Trade Slowdown Based on the Wiod 2016 Release." Groningen Growth; Development Centre, University of Groningen.

Ueno, Asako. 2013. "Multilateralising Regionalism on Government Procurement." OECD Publishing.

Worldbank. 2017. "Benchmarking Public Procurement 2017." Washington DC: World Bank Group.

Yotov, Yoto V, Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. 2016. *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. World Trade Organization Geneva.

3.A World Input Output Database

This appendix contains additional detail on sample selection and variable construction for WIOD based variables, plus some more data descriptive statistics. For the analysis, inter-country flows of public consumption final demand gross and value added imports are constructed from WIOD data. There are $m = 56$ industries and $n = 43$ countries in the database.[12] WIOD input output tables are symmetric, i.e. consuming and producing industries are the same. An element in the inter-country input output table is given by x_{ijkl}, implying a flow from from industry i in country k to industry j in country l.

In this paper, we are interested in inter-country flows only. Construction of gross flows is straightforward by summing the country-industry origin of public final consumption demand over industries by country of origin. Value added flows are constructed adapting the Koopman, Wang, and Wei (2014) methodology. In particular, country-industry value added content of public final demand $VA(FD_{gov})$ is computed by

$$VA(FD_{gov}) = \hat{V} \cdot B \cdot FD_{gov}$$

where $\hat{V}$ is a $mn \times mn$ diagonal matrix with country-industry value added shares on the main diagonal, $B = (I - A)^{-1}$ is the $mn \times mn$ country-industry Leontief inverse and FD_{gov} is a $mn \times n$ matrix with rows of country-industry flows absorbed by public sector consumption final demand in each country. Each element (ik, l) of $VA(FD_{gov})$ then gives the value added content from country-industry ik in the public sector consumption final demand of country l. Country-to-country value added flows are then obtained similarly to gross flows by summing the country-industry flows over industries by country of origin.

In the remainder of this Subsection, some additional descriptive statistics are presented. Figure 3.5 shows import shares for public consumption final demand for each country, aggregated over all trade partners. It can be seen that foreign value added content is consistently substantially higher than for gross imports, for some countries substantially.

[12] In addition, there is a *Rest of the World* aggregate (ROW). Since it cannot be determined which type of agreement flows involving ROW are covered by, all flows from, to and within ROW are dropped from the analysis.

Figure 3.5: Gross and Value Added Government Import Shares

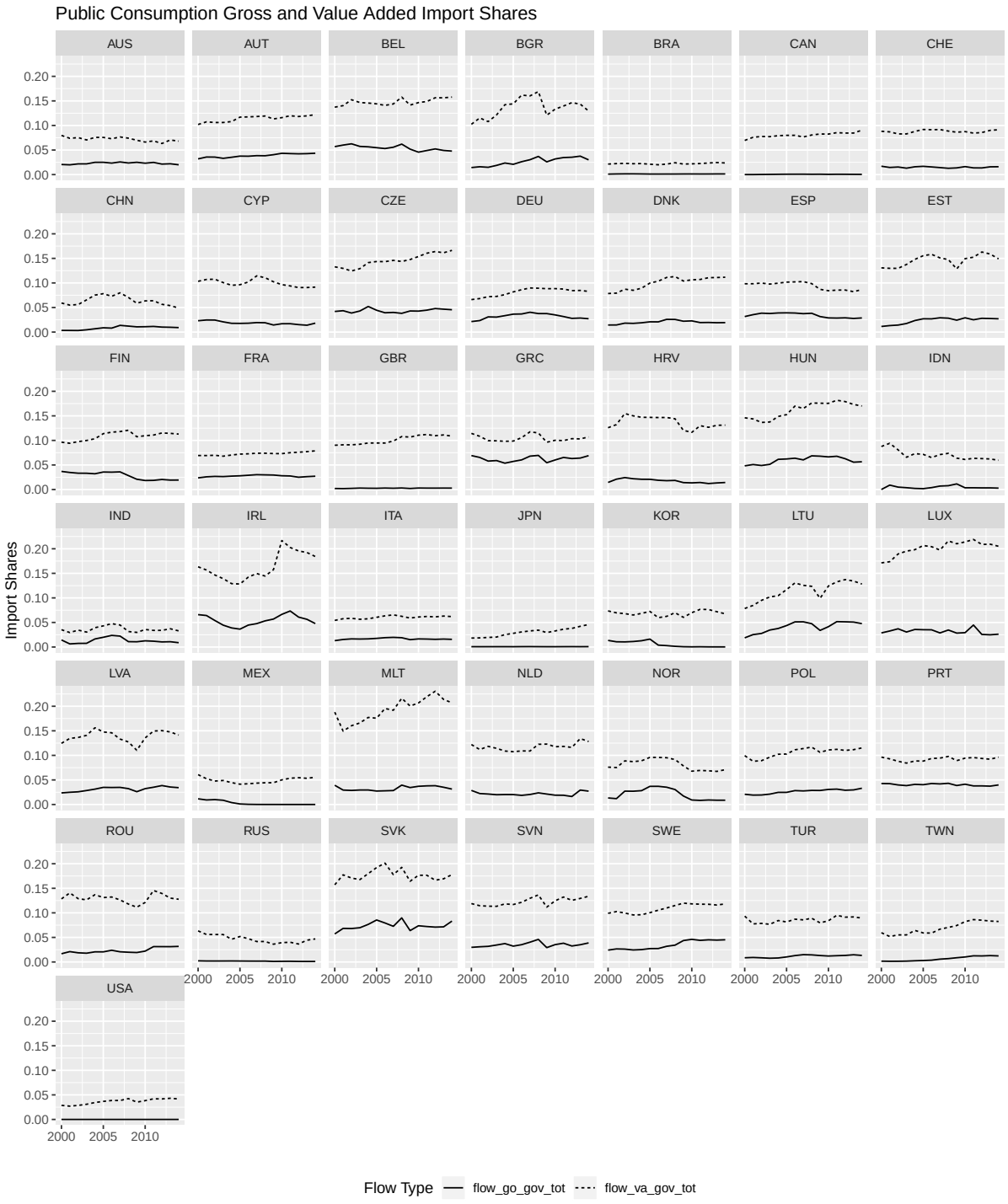
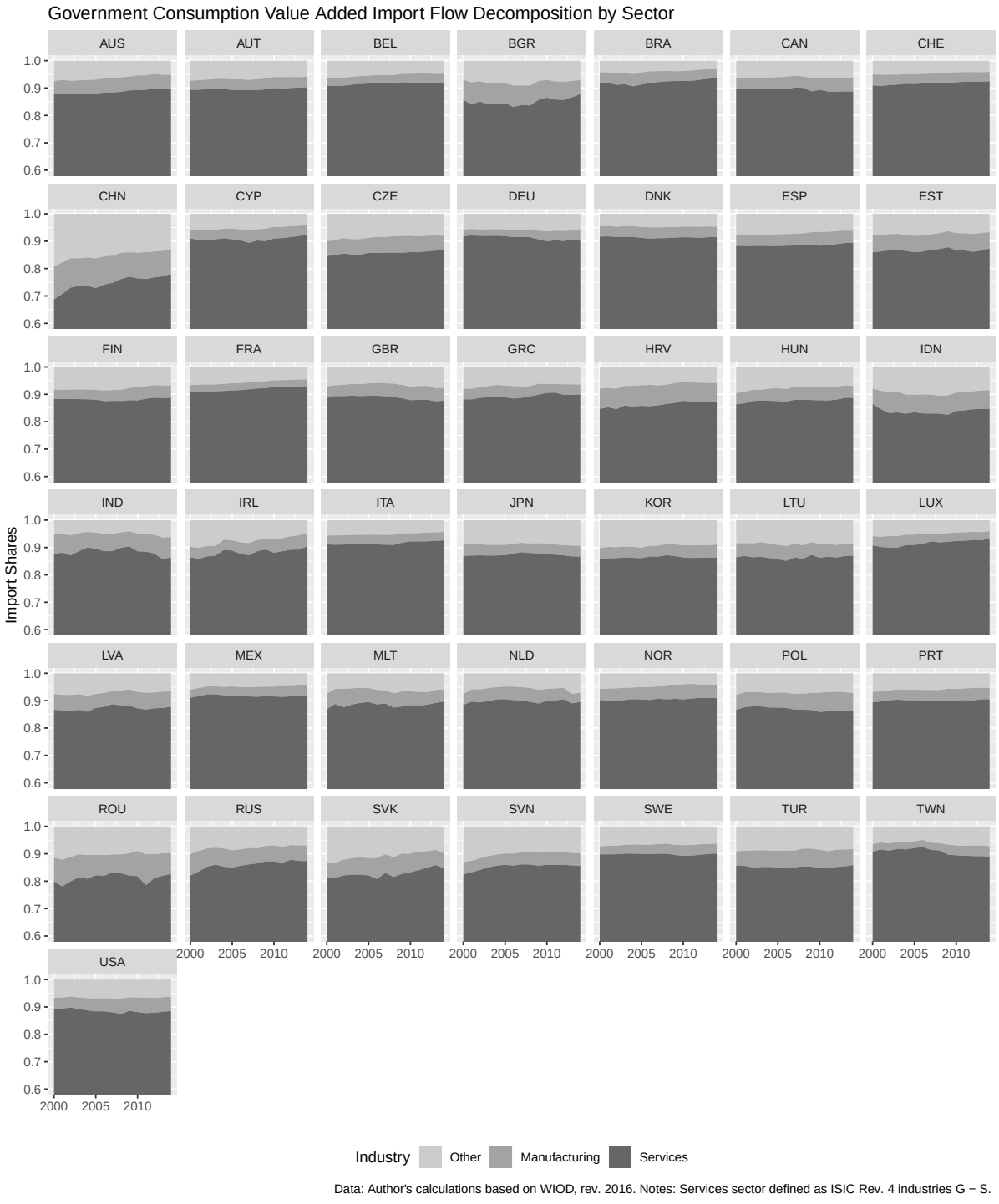


Figure 3.6: Value Added Government Import Flow Decomposition by Sector



3.B Robustness

This appendix contains a number of robustness exercises.

Table 3.11: results-va-gov-tot-2years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2236***				0.0890
	(0.0384)				(0.0765)
provisions		0.2451***		0.0890	0.0005
		(0.0334)		(0.0677)	(0.1010)
enforcable			0.2556***	0.1765*	0.1875*
			(0.0361)	(0.0761)	(0.0773)
N	14792	14792	14792	14792	14792
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

3.B.1 Two-year panel

For comparison with Dhingra, Freeman, and Mavroeidi (2018), whose main results are estimated using 2 year intervals, the baseline results are reestimated on that frequency. The sample includes bilateral trade flows for every second year from 2000 through 2014.

Results are presented in Table 3.11 for total value added imports, in Table 3.12 for value added manufacturing imports and in Table 3.13 for value added services imports. While quantitatively the estimated effects is slightly smaller than in the 4 year interval baseline, the qualitative pattern and significance levels are very similar. These results are consistent with the hypothesis that the economic effects of PTAs require some time to fully materialize. Y. V. Yotov et al. (2016) cite this as the main reason the literature focuses on longer time intervals.

3.B.2 Gross public import flows

The baseline results are presented using the foreign value added content of government consumption expenditures. I argue that this is the preferred way of measuring import flows on theoretical grounds, since it arguably better captures the main modes of PP imports via foreign subsidiaries (c.f. Ramboll and Chur (2011)). Furthermore, the value added measures is net of re-export double-counting flows present in gross trade flows.

Nonetheless, for robustness, results for gross trade flows are presented in Table 3.14 for total public import flows. Like in the value added case, each PTA categorization is strongly significant when tested separately. REsults change when looking at the separate effect

Table 3.12: results-va-gov-man-2years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2609***				0.1402
	(0.0557)				(0.0861)
provisions		0.2850***		0.1960***	0.0571
		(0.0552)		(0.0490)	(0.0964)
enforcable			0.2917***	0.1006	0.1193
			(0.0601)	(0.0708)	(0.0734)
N	14792	14792	14792	14792	14792
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

of shallow and deep provisions. While still jointly significant, neither shallow nor deep provisions are significant individually. Furthermore, the coefficient on shallow provisions is substantially larger. When turning to model (5), only the general agreement coefficient is significant, while the presence and nature of specific PP clauses does not significantly increase the effect (the joint effect is still significantly positive).

When looking at the results for manufacturing and services. separately in Tables 3.15 and 3.16 respectively, results are qualitatively different from the baseline results as well. While for value added imports, the overall results are driven by services, for gross imports both coefficient values and significance levels are higher for manufacturing than for services. For both specifications (4) and (5), enforceability of PP provisions is associated with significancly higher manufacturing import flows to the government sector. In model (5) for the manufacturing sector, agreements with enforceable provisions are associated with significantly higher flows than general PTAs without specific PP provisions. In contrast to the manfucaturing results, when looking at the service sector and gross imports the nature of PP provisions is not associated with significantly higher flows. This again stands in contrast to the value added flows.

3.B.3 Placebo regression

A general drawback of the econometric strategy is the lack of truly exogenous variation. Potentially, in the current context this problem can be substantial. Trade agreements are actively sought and negotiated, making it potentially likely that there will be a selection effect in who signs a (deep) agreement with whom and who does not. To corroborate the

Table 3.13: results-va-gov-ser-2years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2612***				0.1224
	(0.0386)				(0.0627)
provisions		0.2701***		0.0312	−0.0909
		(0.0384)		(0.0928)	(0.1111)
enforcable			0.2843***	0.2579*	0.2678**
			(0.0414)	(0.1009)	(0.1020)
N	14792	14792	14792	14792	14792
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

main findings, the design features of trade agreements allow for a placebo exercise. The idea is to redo the main analysis with an alternative outcome variable that is independent of the explanatory policy variable of interest. If the policy variable truly causally influences the original outcome variable, the effects on the placebo outcome should be insignificant.

The design of trade agreements offers a natrual placebo outcome variable. By definition, PP clauses affect the public sector *exclusively*. This implies that trade flows of private actors should not be affected by the design of PP provisions by definition. For the placebo exercise, I therefore test the effect of PP provisions on non-governmental consumption import flows.

The results of this exercise are presented in Table 3.17 for total bilateral flows, Table 3.18 for manufacturing sector import flows and Table 3.19 for service sector import flows.

Results are qualitatively and quantitatively very similar to the public sector regressions for both aggregate and sector-specific flows. Prima facie this can be seen as evidence against causal interpretatility of the main results. Alternatively, it could also indicate the presence of a confounding factor, casting doubt on the validity of the placebo. In particular, the Shingal and Ereshchenko (2018) classification concerns *exclusively* PP provisions. If 'deeper' PP provisions are more likely to be included in agreements containing 'deeper' general provisions as well, then the results could be reflecting an omitted variable bias, since in the regression deep PP provisions could proxy for deep general provisions. Since Shingal and Ereshchenko (2018) do not categorize PTAs by their general deepness beyond the public sector, this cannot be tested given the currently available data. One conclusion

Table 3.14: results-go-gov-tot-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.3269***				0.2300**
	(0.0534)				(0.0786)
provisions		0.3203***		0.2990	0.0722
		(0.0497)		(0.2138)	(0.2268)
enforcable			0.3206***	0.0295	0.1418
			(0.0510)	(0.2192)	(0.2204)
N	7396	7396	7396	7396	7396
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

however could be that in order to test the strength of PP provisions, potentially it might be necessary to include additional information on the PTAs concerning the general 'depth' of the agreement, also beyond the government sector.

Table 3.15: results-go-gov-man-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.6539***				0.5268***
	(0.1205)				(0.1302)
provisions		0.5658***		0.0302	−0.4959
		(0.1077)		(0.2437)	(0.2715)
enforcable			0.5813***	0.5467*	0.6521*
			(0.1098)	(0.2600)	(0.2674)
N	7388	7388	7388	7388	7388
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

Table 3.16: results-go-gov-ser-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2202***				−0.2523
	(0.0603)				(0.1485)
provisions		0.2345***		0.2471	0.5041
		(0.0570)		(0.2234)	(0.2698)
enforcable			0.2339***	−0.0393	−0.2059
			(0.0586)	(0.2303)	(0.2351)
N	7396	7396	7396	7396	7396
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0004	0.4829

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

Table 3.17: results-va-con-tot-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2030***				0.0007
	(0.0404)				(0.0931)
provisions		0.2388***		0.0400	0.0395
		(0.0284)		(0.0930)	(0.1317)
enforcable			0.2592***	0.2212*	0.2234*
			(0.0276)	(0.0961)	(0.0968)
N	7396	7396	7396	7396	7396
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

Table 3.18: results-va-con-man-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2606***				0.1043
	(0.0411)				(0.0888)
provisions		0.2827***		0.2368***	0.1333
		(0.0319)		(0.0583)	(0.1055)
enforcable			0.2854***	0.0519	0.0671
			(0.0346)	(0.0649)	(0.0660)
N	7396	7396	7396	7396	7396
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

Table 3.19: results-va-con-ser-4years

	Model 1	Model 2	Model 3	Model 4	Model 5
agreement	0.2233***				0.0014
	(0.0398)				(0.0767)
provisions		0.2465***		−0.1093	−0.1102
		(0.0364)		(0.1342)	(0.1544)
enforcable			0.2807***	0.3697**	0.3784**
			(0.0329)	(0.1373)	(0.1378)
N	7396	7396	7396	7396	7396
exporter-time FE	1	1	1	1	1
importer-time FE	1	1	1	1	1
Pair FE	1	1	1	1	1
Joint sign. (p-val)				0.0000	0.0000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered by importer-exporter.

Chapter 4

Determinants of Input Price Dispersion

This Chapter is joint work with Árpád Ábrahám and T. Kirk White. [1]

Abstract

Two recent contributions document that (a) supply chain considerations are not the prime focus of vertical integration and (b) input prices display substantial heterogeneity across U.S. firms. This paper outlines a strategy to empirically test whether vertical integration and input price dispersion are related, using unique features of U.S. Economic Census micro data that have already been employed by Atalay et al. (2014) and Atalay (2014). In particular, we argue that owning productive capacities upstream endows firms with informational or bargaining advantages which result in lower procurement prices. We propose to decompose input price dispersion in a between-supplier and a within-supplier component in order to investigate how both components contribute to the input price advantage of vertically integrated firms. This project aims at shedding light on the determinants of measured firm productivity. Input price dispersion through the between-supplier component predicts measured productivity dispersion through supplier selection, while within-supplier price dispersion implies that measured productivity dispersion is driven by differences in supplier market power.

[1]Some of the research in this paper was conducted while the third author was an employee of Census Bureau. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. We would like to thank Robert Goodhead for valuable comments.

4.1 Introduction

Motivation The majority of the economics literature on Vertical Integration (VI) is rationalizing integration decisions with supply chain considerations. However, Atalay, Hortaçsu, and Syverson (2014) present evidence that supply chain considerations are *not* the primary reason for the existence of vertically integrated firms, using establishment- and shipment-level data of the Census of Manufactures (CM) and the Commodity Flow Survey (CFS). In fact, more than half of all vertically integrated establishments feature *zero* intra-firm shipments (see Fig. 1 below, reproducing their main result).

In a related paper, Atalay (2014) shows that even in industries characterized by commodity-like products, there is substantial input price dispersion. He focuses on the consequences of accounting for both plant level input and output prices in the estimation of plant productivity. Importantly, Atalay (2014) relaxes perfect competition in input markets by allowing firm specific prices. However, he assumes that firms take input prices as given, i.e. input price dispersion is assumed to be exogenous and firms do not optimize on how to procure inputs.

In this project we propose to test empirically whether vertical integration decision and input price dispersion are related using the CM and CFS data employed by Atalay (2014) and Atalay et al. (2014).[2] In particular, we aim at testing whether the VI status has an indirect effect on procurement outcomes even in the case where an upstream establishment is not acting as a direct supplier to customer plants of the firm. We argue that owning productive capacities upstream endows firms with informational or bargaining advantages which result in lower procurement prices when bargaining with suppliers outside the firm. This could offer a potential explanation for the observed lack of intra-firm physical good shipments documented by Atalay et al. (2014), by linking input price dispersion to the vertical integration status of a firm.

A better understanding of the economic mechanisms underlying measured productivity differences matters for economic policy. In particular, the framework predicts that observed productivity differences can be driven either by firms procurement efficiency or the exploitation of non-competitive pricing. Which effect dominates has implications for the interpretation of observed productivity, namely whether it is a sign of high efficiency or high market power.

In addition, by combining data from the CM and the CFS, this project enables an evaluation of the imputation methodology used to impute missing values in the CFS data.

[2]We will elaborate more on how our data construction differs and expands on the data employed in those papers in Section 2.

Figure 4.1: Value Share of Internal Shipments

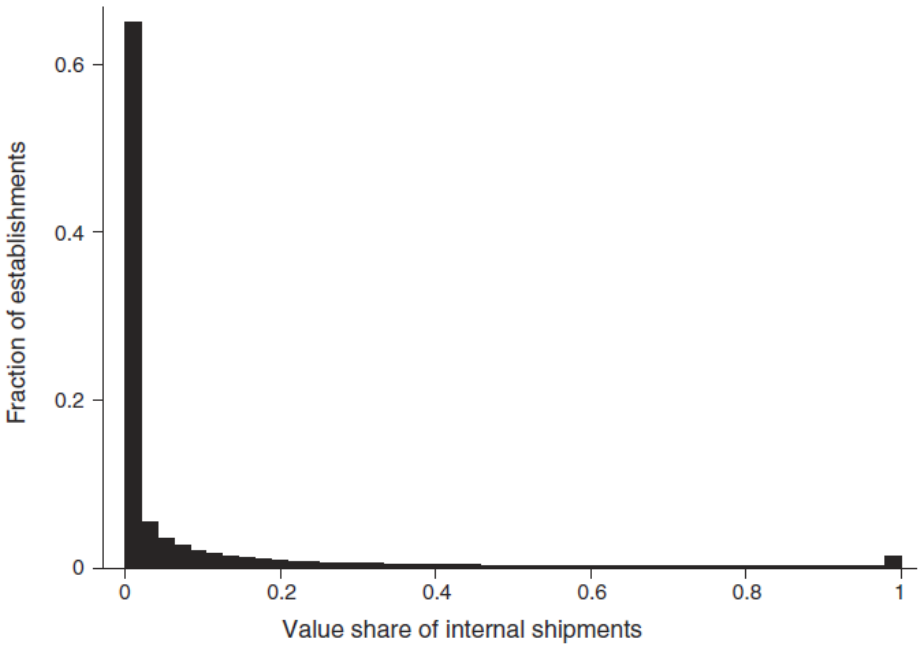


FIGURE 1. SHARE OF INTRAFIRM SHIPMENTS
BY UPSTREAM VERTICALLY INTEGRATED ESTABLISHMENTS

Source: Atalay, Hortaçsu, and Syverson (2014), Figure 1.

This project is therefore also useful in ensuring the data quality of U.S. Census Bureau micro data, in particular the Commodity Flow Survey data.

Mechanism Potentially, VI can influence input prices in two ways. First, acquiring a subsidiary in a supplier industry gives a downstream firm access to detailed knowledge about the upstream industry's competitive situation, input cost structure and productivity.[3] The firm can potentially use this knowledge to target the most efficient upstream producers in the process of procuring inputs from the respective industry.[4]

Secondly, VI potentially enhances the firm's bargaining position with potential input suppliers. Owning productive capacity delivers an outside option for procurement, available without delay by avoiding the need to enter additional bargaining with alternative suppliers.

[3] Mueller et al. (2016) provide survey evidence from the German automotive industry that downstream firms consider the informational implications of outsourcing important. In particular, they cite managers stating that the loss of knowledge following the outsourcing of productive capacity could not be adequately replaced by hiring consultancy knowledge.

[4] For example, bidding for procurement contracts of other buyers through the upstream subsidiary will give the firm information on the identity of the winning upstream firm, which is likely to be indicative of it's efficiency.

The decomposition of input price dispersion into a within- and between-supplier component enables a tentative test of these two channels. A bargaining advantage would imply that a vertically integrated customer can procure at a lower cost compared to another non-integrated customer of the same supplier. An informational advantage would imply that a vertically integrated firm is able to target cheaper suppliers compared to non-integrated firms, due to better knowledge of the market in question.

Arguably, the informational advantage of being vertically integrated can also lead to bargaining advantages in itself, e.g. if bargaining entails an informational asymmetry. Importantly though, only the information channel would predict that VI implies advantages at targeting potential suppliers relative to non-integrated firms. The bargaining channel works exclusively through the price determination with a particular supplier.

Both the information and the bargaining channel predict vertically integrated establishments to procure at lower input prices. However, they predict different mechanisms: The information channel predicts that integrated firms procure from more efficient suppliers. It remains silent on within-supplier price dispersion. The bargaining channel on the other hand predicts that procurement price advantages are achieved through getting better prices from the same suppliers than non-integrated firms. Put differently, the information channel predicts price advantages through the *extensive* margin, while the bargaining channel predicts price advantages through the *intensive* margin. Thus, bargaining advantages have implications for within-supplier price spreads, while the information channel predicts price differences between suppliers and relative to the industry average price. The U.S. Economic Census data described below offer a unique way to distinguish between these two mechanisms empirically.

Predictions The mechanism just sketched has the following two main implications for between- and within-supplier price dispersion:

(i) *Within-industry price dispersion:* Everything else equal, firms which are vertically integrated into industry j procure inputs from j at a lower price than non-integrated firms. This is true for both the information and the bargaining channel.

(ii) *Between-supplier price dispersion:* Vertically integrated firms should be able to target cheaper suppliers. Therefore, they should procur from producers that are on average cheaper than non-integrated firms procuring from the same industry. This prediciton is unique to the information channel.

(iii) *Within-supplier price dispersion:* Suppliers charge customers with more upstream links and upstream productive capacity relatively lower prices compared to customers without upstream establishments. This prediction is unique to the bargaining channel.

In addition, there are auxiliary predictions of the information channel indirectly related to price dispersion:

(iv) Through better information, vertically integrated establishments are linked to fewer upstream sellers, as the firm is able to target potential suppliers more precisely. This is an implication of the information channel.

(v) Better targeting also implies that the firm, on average, procures from more cost-efficient plants relative to other firms without a subsidiary in the respective industry.

In order to test these predictions empirically, establishment-level data on input prices is necessary. Given these, predictions (i) to (iv) can be tested empirically by combining input price level data with data on the following items:

1. A plant-level indicator of vertical integration for each buyer-supplier relationship. More precisely, such an indicator would be equal to one, if the buying plant is part of a firm owning upstream capacity in the industry of the supplying plant and zero otherwise.

2. A variable indicating the number of upstream suppliers of a particular establishment from a particular industry.

3. A variable for the cost efficiency/productivity of the supplying plant relative to the average productivity in the respective industry.

4. A variable capturing the relative price a firm pays relative to other customers of the same supplier.

The empirical strategy is described in more detail below, following a discussion of the U.S. Economic Census data used to construct items 1 – 4.

4.2 Data

The hypotheses can be tested using establishment and shipment level data from the U.S. Economic Census. Similar data have already been constructed and used by Atalay et al. (2014) and Atalay (2014) for their analyses. The main data sources are the Census of Manufactures (CM) and its Materials and Production Supplement and the Commodity Flow Survey (CFS) at establishment and shipment level of detail, respectively. The combination of CM and CFS data enables the identification of buyer-supplier pairs in combination with price and quantity information on shipments, both of which is necessary to test the maintained hypotheses and are a unique to the Economic Census data. In addition, we draw on data from the Longitudinal Business Database (LBD), the BEA Input-Output tables and the NBER-CES Manufacturing Productivity Database.

CM The CM contains variables on establishments productive capacities such as employment, wage bill, book value of capital and structures as well as electricity expenditures. In addition, a subset of industries and establishments provide quantity and value of products produced (at seven-digit level) and consumed (at six-digit level). The CM is conducted in years ending in '2' and '7'. While Atalay (2014) focuses on a small subset of industries with homogeneous output, Atalay et al. (2014) use the establishments in the Longitudinal Business Database, spanning a much wider range of industries. Since it will be possible to control for industry effects, for this project we would focus on the latter, more complete sample.

CFS The CFS is part of the Economic Census and provides information on the type, destination, origin, weight, value and mode of transportation of commodities shipped. The CFS covers mining, manufacturing, wholesale and some service trade industries in the U.S. By providing precise information on source and destination of shipments, the CFS indirectly enables the identification buyer-supplier establishment pairs. For more details, see below.

Data construction Constructing the data necessary to test the hypothesis on the relationship of VI and input price dispersion poses two main challenges. First, we need an indicator for what is a vertically integrated plant and second we need to identify both the sender and the recipient plant of a particular shipment and the latter's firm. Constructing those measures closely follows the work done by Atalay et al. (2014) and Atalay (2014). While they analyze the data based on the characteristics of the supplying plant, we are more interested in the perspective of the consuming plant. The variable definitions are very closely related nonetheless. Atalay et al. (2014) are interested exclusively in shipments of firms that are vertically integrated, whereas we look at the entire sample of shipments, including shipments by non-integrated firms. Furthermore, they do not exploit the information on shipment prices available from the CFS. Atalay (2014) employs the shipment price information from the CFS to control for plant level input and output price dispersion. His analysis is limited to a small number of industries, characterized by homogeneous products. We are looking at price dispersion in the entire manufacturing sector. In addition to the VI indicator, we need to construct a measure of plant productivity and create measures of price dispersion. We will discuss the different parts in more detail next.

1. Vertical integration indicators for establishments

The final aim is to construct vertical integration indicators on shipment level, which are the main variables of interest for our analysis, as we try to distinguish procurement

outcomes of vertically integrated versus non-integrated firms. Shipments are flagged as vertically integrated if the recipient plant is vertically integrated into the sending plant's industry. The creation of VI indicators for establishments requires two parts. First, it is necessary to map the relative upstreamness of industries and second we need to identify whether a plant is part of a firm with productive capacity in the respective upstream industry. For the second step, we also need to identify the plant receiving the shipment, which is done in Step 3. below.

In order to map the relative upstreamness of industries make use of information from the U.S. Input-Output tables. In particular, two industries i and j are defined as vertically linked if upstream industry j ships more than 5% of it's total ouput to industry i.[5] Information on the split of sales across industries is obtained from the BEA's Input-Output tables.[6]

Second, using this industry level information, a downstream establishment in an industry i is identified as vertically integrated into an industry j if it is owned by a firm that owns a plant in an industry j located upstream in the product space as well. The necessary plant ownership indicators are obtained from the Longitudinal Business Database (LBD) data.[7] Finally, as already mentioned above, shipments are flagged as vertically integrated if the recipient plant is vertically integrated into the sending plant's industry.

2. Sender and receiver establishment identities of shipments

In order to connect buyer-supplier relationships with other establishment characteristics, we need to identify both ends of each transaction. Sender establishment identities are directly available from CFS data. However, this is not the case for receiver establishments. Instead, the CFS data contains only zip-code destinations of shipments. Therefore, the recipient plant of a shipment has to be imputed. In order to do this, we again closely follow the approach taken by Atalay et al. (2014). Potentially, any plant located in the destination zip code of a shipment could be the receiving entity. Among all plants in the

[5]This cutoff could be varied for robustness. Generally, a lower cutoff will make more shipments vertically integrated, i.e. expand our overall sample. This comes with the caveat that with a low cutoff we include additional industry pairs in the analysis between which shipments are not frequent. As long as price dispersion along those links is not different from higher volume links, this should not present a problem.

[6] Atalay et al. (2014) use additional, more disaggregated information from wholesale and retail trade sector surveys in order to achieve a finer disaggregation of industries and to trace shipments mediated by retailers/wholesalers. For simplicity, we focus on the readily available Input-Output information.

[7]Our definition of VI status differs slightly from Atalay et al. (2014). While they are interested in the VI status of *upstream* plant's owner, for the bargaining perspective we are interested in, the integration of the *downstream* plant's firm is the relevant dimension.

destination zip code of a shipment, we first keep only those that belong to an industry downstream of the sending plant's industry. If among the remaining candidate plants, there is a plant belonging to the same firm as the sending plant, we assign this plant as recipient. This implies that if possible, the shipment will be counted as taking place within the firm. This choice reflects our prior that shipments within the firm are more likely than shipments between firms.[8] If there are still multiple remaining recipient candidates left, we assign one of the remaining plants randomly.[9]

3. Marginal cost/productivity of suppliers

The information channel predicts procurement of integrated firms from relatively more efficient suppliers. In order to control for the efficiency of suppliers, we construct standard revenue productivity measures using industry level price deflators and industry level factor shares coming from the NBER-CES Manufacturing Industry Database. For calculating productivity, we assume that plants within an industry produce using a gross output Cobb-Douglas production function with constant returns to scale and common factor shares including capital, labor, materials and energy as production factors.

4. Within-industry, between-supplier and within-supplier price spreads

The definition of average price spreads for this project is different from the definition used by Atalay (2014). The reason is that we are interested in the relationship between VI and the degree of price dispersion. Since VI is defined only with respect to a particular upstream industry, the measures of price dispersion have to be industry-specific as well. We will define measures of within-industry, between-supplier and within-supplier price dispersion in turn, which are used to test hypotheses (i) to (iii). To fix ideas, lets define a particular shipment s, the supplying (upstream) industry I, with plants $i \in \{1, ..., I\}$, the buying (downstream) industry J, with plants $j \in \{1, ..., J\}$ and the entire universe of plants N, with $n \in \{1, ..., N\}$.

Define the shipment price of a shipment from plant i to plant j in year t as

$$P_{sijt} \equiv \frac{\omega_{sijt}}{\chi_{sijt}},$$

[8]Shipments assigned as internal will be dropped in the subsequent analysis, see also detailed discussion of sample selection below.

[9]Further robustness checks on the assignment algorithm are possible. A first one would be to weigh the assignment probability by the size of the plants, or by the amount of materials consumed. An additional robustness check would be to exploit information on the type of good shipped: From the CM data, we know the main type of material consumed by a plant. This can potentially be matched against the commodity code of the shipment. This requires a concordance between CM product code classification and the STCC product classification used in the CFS, which is work in progress.

where χ_{sijt} is the weight of the shipment and ω_{sijt} the shipment value. From now on, drop t for expositional clarity. We follow Atalay et al. (2014) and aggregate all shipments between two plants in a year. To do that, we define the average shipment price from plant i to plant j in year t as

$$P_{ijt} \equiv \frac{\omega_{ijt}}{\chi_{ijt}} = \frac{\sum_{sijt} \omega_{sijt}}{\sum_{sijt} \chi_{sijt}},$$

where χ_{ijt} is the total mass of shipments in year t and ω_{ijt} the total value in the respective year. Both are available from the CFS by aggregating quantities and values from all individual shipments s from i to j in t.

Within-industry price dispersion. This measures whether a firm $j \in J$ pays less for inputs from an upstream industry I than what the upstream industry charges on average. Start by defining the average value-weighted price charged by I, which is given by

$$\overline{P}_I = \sum_{i \in I} \frac{\sum_{n \in N} \omega_{in} P_{in}}{\sum_{n \in N} \omega_{in}}.$$

The firm specific average price charged by a firm $j \in J$ from industry I is given by

$$\overline{P}_{Ij} = \frac{\sum_{i \in \Gamma_I(j)} \omega_{ij} P_{ij}}{\sum_{i \in \Gamma_I(j)} \omega_{ij}},$$

where $\Gamma_I(j)$ denotes the set of suppliers of firm j from industry I. The measure of within-industry price dispersion is then defined as

$$\overline{\psi}_{Ij} \equiv \log\left(\frac{\overline{P}_{Ij}}{\overline{P}_I}\right).$$

Between-supplier price dispersion. Prediction (ii) states that a vertically integrated firm will on average procure from lower cost producers than non-integrated firms. For this, we need to compare the average price paid for inputs from each supplier i from industry I by firm j, to the average price of that supplier. Begin by defining the average price of a firm i from industry I. It is given by

$$\overline{P}_i = \frac{\sum_{n \in N} w_{in} P_{in}}{\sum_{n \in N} w_{in}}.$$

The relative price of this firm wrt it's own industry is defined as $\overline{\psi}_{\bar{i}} \equiv \log(\overline{P}_i / \overline{P}_I)$. Using this, the relative price of downstream firm j's suppliers from supplier industry I is given by

$$\overline{\psi}_{\bar{i}j} = \frac{\sum_{i \in \Gamma_I(j)} w_{ij} \overline{\psi}_{\bar{i}}}{\sum_{i \in \Gamma_I(j)} w_{ij}}$$

Within-supplier price dispersion. Prediction (iii) states that given a supplier i, a VI-establishment j pays less than a non-integrated establishment j'. Thus, we are

interested in the price charged by i from this particular firm, P_{ij}, relative to the average price charged by the supplier i, $\overline{P}_i$, define this as $\psi_{ij} \equiv \log(P_{ij}/\overline{P}_i)$. Then, averaging over all suppliers of j from industry I, we get

$$\overline{\psi}_{ij} = \frac{\sum_{i \in \Gamma_I(j)} w_{ij} \psi_{ij}}{\sum_{i \in \Gamma_I(j)} w_{ij}}.$$

Sample Selection Our sample covers establishments that are present in the CFS, the CM and the LBD. The CFS covers mining, manufacturing, retail and wholesale trade industries. We restrict the sample to plants present in the Census of Manufactures, since we use that information to construct productivity and other plant measures for controls. We drop all shipments originating from or go to so called Administrative Records (AR) establishments from the Census, since these are small establishments (up to 5 employees) for which almost all data are imputed. This strategy follows the typical approach in the literature. We further restrict the sample to domestic shipments and drop all shipments destined for export. We also drop shipments between two plants owned by the same firm. The prices in our data are in fact unit values of shipments. Our maintained assumption is that for shipments between firms, the value of a shipment reflects the price paid by the buyer. Since shipments within firms are not necessarily priced at market value, it is unclear what the reported shipment values actually reflect in those cases. For this reason, we restrict the analysis to shipments between different firms.

Decomposition of Price Dispersion We will be particularly interested in shipment characteristics by sending plant, the relative frequency of internal, vertically integrated and non-integrated shipments across firms of different sizes as well as the decomposition of overall price dispersion into the between-supplier and the within-supplier components for the full sample across shipments. Since our dispersion measures are in logarithms, the total price deviation of a particular buyer-supplier link can be decomposed as

$$\underbrace{\log(\frac{P_{link}}{P_{ind}})}_{\text{overall dispersion}} = \underbrace{\log(\frac{P_{link}}{P_{sender}})}_{\text{within sender dispersion}} + \underbrace{\log(\frac{P_{sender}}{P_{ind}})}_{\text{between sender dispersion}} .$$

This distinction can help to assess the economic sources of observed price dispersion. A large share of between price dispersion hints at the presence of heterogeneous firms in the supplier market, while a large share of within price dispersion indicates that firms are able to price discriminate across different customers. In a perfectly competitive market, both within and between sender price dispersion should be zero.

Sample including Recipient Output Prices The CFS is recorded only at the sender level. However, in our recipient assignment algorithm we choose plants as recipients from

the same underlying population of the CMF that the CFS was drawn from. Therefore, there is a positive probability that a plant assigned as recipient in one particular shipment is sampled in the CFS as a sender for another shipment. This implies that for a subsample of the shipments, we will know the average prices charged for the output of the recipient plant. This is of interest in our context since it enables us to partially control for product quality when estimating the effect of VI on procurement prices.

4.3 Estimation Strategy

The basic exercise for testing hypotheses (i) to (iii) is to regress industry price deviation, between-supplier price deviation and within-supplier price deviation of a buyer-supplier link on the indicator for vertical integration of the buying plant. The baseline specification is given by

$$x_{ijt} = \beta_0 + \beta_1 \mathbf{1}_{\{VI_j(i) \neq \emptyset\}} + \beta_C C_{hit} + \epsilon_{hit}.$$

Here, $VI_j(i)$ is the set of plants in the industry of i of the firm owning j and $\mathbf{1}$ the indicator function. C_{hit} contains a vector of controls, such as firm size and industry and time fixed effects. The set of controls will be discussed in more detail below. The dependent variable x_{ijt} is chosen depending on which hypothesis (i) to (v) is tested. For the main hypotheses (i) to (iii), of within-industry, between-supplier and within-supplier price dispersion, x_{sijt} is given by $\overline{\psi}_{Ij}$, $\overline{\psi}_{\bar{i}j}$ and $\overline{\psi}_{ij}$, respectively. It is important to note that in order to test hypotheses (i) to (v), we are interested in the aggregate relationship between input prices and VI and not in the characteristics of individual firms. Therefore, the hypotheses can be tested and presented while ensuring the confidentiality of individual firm characteristics.

With this empirical specification, we aim to estimate the effect of VI on the price of a shipment, relative to the sender or industry average price. The main challenge for our estimation is the fact that in principle the VI status of a plant is a choice variable of the owning firm and that we do not have a truly exogenous source of variation for the VI status of a particular plant. Our approach is therefore to match firms as closely as possible with exception of VI status and control for differences between plants that might be driving VI status and/or the relative price of the transaction.[10]

We control for three major types of confounders: firm size, plant 'type' and product quality. Regarding firm size, larger firms own more plants and are therefore mechanically more likely to be vertically integrated. We therefore want to compare plants owned by firms

[10]While we cannot rule out reverse causality, we think that in our setup it is not likely to drive the results. For relative input prices to impact the VI status, non-integrated firms facing relatively high prices would have to buy a plant in the suppling industry with the aim of getting better prices from their (now) competitors, rather than with the aim of producing their own inputs. This seems like a rather unlikely way to improve on procurement costs.

of similar size. Second, we control for plant type by controlling for plant size, TFP and capital-labor ratio. Plant size directly influences the quantities bought, which will have an impact on procurement costs as long as pricing is non-linear. TFP and K/L might be indicators of product quality. As a more direct control of product quality, we will control for buyers' output prices. This will potentially come at the cost of reducing the sample size.

An additional, clearly important control is shipment distance. The shipment prices are free-on-board prices excluding shipment costs, mitigating the importance of this control, nonetheless, it is a natural candidate. The construction of this is work in progress.[11]

In addition to those controls, we include year fixed effects and a set of industry fixed effects. In particular, we exploit variation between senders of the same industry, between recipients of the same industry and to particular input-output industry links.

Alternative estimation strategy The main estimation strategy just outlined relies on cross-sectional variation in VI. We choose this strategy closely following the empirical setup of Atalay (2014). In order to corroborate the evidence, an alternative approach would be to exploit the panel dimension of the data, since the CFS is available in five-year waves. In particular, given the data it is in principle possible to estimate how procurement prices change following a change of establishment ownership between firms with different VI status. The analysis in Atalay et al. (2014) indicates that there is a substantial number of changes in VI status among the plants in the Census of Manufacturers.[12] Thus, an alternative estimation strategy could rely on changes in procurement prices following a change in vertical integration status. However, the feasibility of such an analysis crucially depends (i) on the number of establishments in the CFS which are present throughout subsequent survey waves and (ii) on the extend to which these data coincide with changes in the VI status in the CM data.

In addition, the particular nature of our data implies that standard panel methods cannot be directly applied. The panel variable in our analysis would be the sender and/or recipient plant. However, since the underlying data are on shipment level, plant identity does not uniquely identify an observation within a year. This is also true when aggregating

[11] In addition, it might be useful to look at the relative distance between the recipient plant and the shipping plant, and the recipient plant and the plant making a shipment vertically integrated. This might in particular play a role in determining the choice between VI and internal shipments, which is not the direct subject of our analysis.

[12] Table 3, Panel D in Atalay et al. (2014) indicates that out of their CM sample observations, about 5% are changing ownership status from non-VI to VI. For our analysis, considering in addition the reverse change from VI to non-VI would further generate additional observations.

shipments between two particular plants.

Evaluating the CFS Imputation In addition to shedding light on the determinants of observed firm productivity, this project can serve to evaluate the imputation procedure of the CFS. As noted in the previous paragraph, the CFS data are partially imputed. Importantly, these imputations do not differentiate between different shipment receivers.[13] Thus, the imputation procedure implicitly assumes that value and quantity of shipments do not vary systematically across recipients. In contrast, the hypothesis we propose to test predicts exactly that shipments prices (i.e. value-quantity ratios) vary systematically with the identity of the receiving establishment, depending on whether it is part of a vertically integrated firm or not. By merging the CM and the CFS data, this project potentially adds information to the recipient establishments of a shipment in the CFS data. We propose to use this in order to evaluate whether the implicit assumption made in the imputation is valid. Thereby this project potentially contributes to the quality control of U.S. Census Bureau data, in particular the CFS data.

4.4 Concluding Remarks

In this project we outline a strategy to test whether input price dispersion and vertical integration status of firms are linked. In particular, we argue that owning productive capacities enables firms to exploit informational and bargaining advantages in order to achieve lower procurement prices. We describe how the specific features of U.S. Economic Census data from the Census of Manufacturers and the Commodity Flow Survey, in particular rare information on input and output prices, can be used to separately identify and test the information and bargaining channel. This data has already been used by Atalay (2014) and Atalay et al. (2014) to establish the stylized facts that motivate the question of this paper, however, they focus on VI and input price dispersion in isolation and do not link the two phenomena.

This project aims at shedding light on the determinants of firm productivity. Given the importance of firm productivity for aggregate economic performance, it is important to identify the economic mechanisms driving differences in observed plant productivity. By linking plant productivity to firm organization, the present proposal also aims at contributing to the understanding of the effects of mergers and acquisitions. In particular, this we suggest to test a mechanism for potential gains from ownership integration, namely

[13] According to the U.S. Census Bureau documentation on the 2012 CFS (http://www.census.gov/econ/cfs/2012_methodology.html), imputations are done using a donor imputation model. Donor shipments are identified using information on sender identity and shipment class, but do not include information on the destination of the shipment.

advantages in input procurement.

In addition, by combining data from the CM and the CFS, this project enables an evaluation of the imputation methodology used to impute missing values in the CFS data. This project is therefore also useful in ensuring the data quality of U.S. Census Bureau micro data, in particular the Commodity Flow Survey data.

Bibliography

Anderson, J. E. and E. Van Wincoop (2003). Gravity with gravitas: a solution to the border puzzle. *American economic review 93*(1), 170–192.

Anderson, R., A. Muller, O. Kodjo, J. P. De Leon, and P. Pelletier (2011). Government procurement provisions in regional trade agreements: A stepping stone to gpa accession. chapter 20 in the wto regime on government procurement: Recent developments and challenges ahead.

Anderson, R., C. M"uller, K. Osei-Lah, J. P. de Leon, and P. Pelletier (2011). Government procurement provisions in regional trade agreements: a stepping stone to gpa accession? In S. Arrowsmith and R. Anderson (Eds.), *The WTO Regime on Government Procurement: Challenge and Reform*. Cambridge University Press.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics 132*(4), 1553–1592.

Atalay, E. (2014). Materials prices and productivity. *Journal of the European Economic Association 12*(3), 575–611.

Atalay, E., A. Hortaçsu, and C. Syverson (2014). Vertical integration and input flows. *The American Economic Review 104*(4), 1120–1148.

Autor, D. et al. (2010). The polarization of job opportunities in the us labor market: Implications for employment and earnings. *Center for American Progress and The Hamilton Project*.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *The American Economic Review 103*(5), 1553–1597.

Bagwell, K. and R. Staiger (2002). *The Economics of the World Trading System*. Cambridge, MA: MIT Press.

Baier, S. L., J. H. Bergstrand, and M. Feng (2014). Economic integration agreements and the margins of international trade. *Journal of International Economics 93*(2), 339–350.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *The Journal of Political Economy 75*(4), 352–365.

Berman, E., J. Bound, and S. Machin (1998). Implications of skill-biased technological change: international evidence. *The quarterly journal of economics 113*(4), 1245–1279.

Bovenberg, A. L. and B. Jacobs (2005). Redistribution and education subsidies are Siamese twins. *Journal of Public Economics 89*(11), 2005–2035.

Bowlus, A. J. and C. Robinson (2012). Human capital prices, productivity, and growth. *The American Economic Review 102*(7), 3483–3515.

Breton, A. and P. Salmon (1996). Are discriminatory procurement policies motivated by protectionism? *Kyklos 49*(1), 47–68.

Cantore, C. M. and S. Togan (2017). Public procurement in the eu. In B. M. H. Aris Georgopoulos and P. C. Mavroidis (Eds.), *The Internationalization of Government Procurement Regulation*, Chapter 5. Oxford: Oxford University Press.

Carballo, J., K. Handley, and N. Limão (2018). "trade cold wars and the value of agreements during crises". Technical report, VoxEU.

Cernat, L. and Z. Kutlina-Dimitrova (2016). Ttip and public procurement: Going beyond the tip of the iceberg. Technical report, CEPS Policy Brief, No. 339.

Ciccone, A. and G. Peri (2005). Long-run substitutability between more and less educated workers: evidence from us states, 1950–1990. *The Review of Economics and Statistics 87*(4), 652–663.

Coen-Pirani, D. (2017). Geographic mobility and redistribution: A macro-economic analysis.

Cortes, G. M., N. Jaimovich, and H. E. Siu (2017). Disappearing routine jobs: Who, how, and why? *Journal of Monetary Economics 91*, 69–87.

Deardorff, A. (1987). Safeguards policy and the conservative social welfare function. In H. Kierzkowski (Ed.), *Protection and Competition in International Trade.* Oxford: Basil Blackwell.

Dengler, B. and B. Hoekman (2018). The wto government procurement agreement as a commitment device: A first economic appraisal. mimeo.

Dhingra, S., R. Freeman, and E. Mavroeidi (2018). Beyond tariff reductions: What extra boost from trade agreement provisions. mimeo.

Djankov, S., A. Islam, and F. Saliola (2016). "how large is public procurement in developing countries?". Technical report, Peterson Institute for International Economics: Washington DC.

Dür, A., L. Baccini, and M. Elsig (2014). The design of international trade agreements: Introducing a new dataset. *The Review of International Organizations 9*(3), 353–375.

Dwyer, R. E. (2013). The care economy? gender, economic restructuring, and job polarization in the us labor market. *American Sociological Review 78*(3), 390–416.

Ethier, W. (2007). 'the theory of trade policy and trade agreements: A critique'. *European Journal of Political Economy 23*, 605–623.

Evenett, S. and A. Shingal (2006). Monitoring implementation: Japan and the wto agreement on government procurement. In S. Evenett and BHoekman (Eds.), *Economic*

Development and Multilateral Trade Cooperation. London: London: Palgrave-McMillan.

Evenett, S. and A. Shingal (2016). Gauging procurement policy change during the crisis-era: Evidence from the global trade alert. PRONTO working paper, Bern: World Trade Institute.

Evenett, S. J. and B. M. Hoekman (2005). Government procurement: market access, transparency, and multilateral trade rules. *European Journal of Political Economy 1*(21), 163–183.

Francois, J. and W. Martin (2004). Commercial policy, bindings and market access. *European Economic Review 48*, 665–679.

Fronk, J. (2015). *International Agreements on Trade in Government Procurement: Formation and Effect*. Ph. D. thesis, Ph.D Thesis submitted to the Faculty of the Graduate School of Arts and Sciences of Georgetown University.

Goos, M., A. Manning, and A. Salomons (2014). Explaining job polarization: Routine-biased technological change and offshoring. *The American Economic Review 104*(8), 2509–2526.

Gourdon, J. and J. Messent (2017). Emerging policy issues: Measures affecting trade in government procurement process. Technical report, Paris: OECD. Trade and Agricultural Directorate Trade Committee Working Party of the Trade Committee (TAD/TC/WP (2015)26/FINAL.

Grossman, G. (2016). "the purpose of trade agreements". In K. Bagwell and R. Staiger (Eds.), *Handbook of Commercial Policy Volume 1A*. Elsevier.

Guvenen, F. and B. Kuruscu (2010). A quantitative analysis of the evolution of the us wage distribution, 1970–2000. *NBER Macroeconomics Annual 24*(1), 227–276.

Guvenen, F. and B. Kuruscu (2012). Understanding the evolution of the us wage distribution: A theoretical analysis. *Journal of the European Economic Association 10*(3), 482–517.

Guvenen, F., B. Kuruscu, and S. Ozkan (2014). Taxation of human capital and wage inequality: A cross-country analysis. *The Review of Economic Studies 81*(2), 818–850.

Handley, K. (2014). Exporting under trade policy uncertainty: Theory and evidence. *Journal of International Economics 94*(1), 50–66.

Handley, K. and N. Limão (2015). Trade and investment under policy uncertainty: Theory and firm evidence. *American Economic Journal: Economic Policy 7*(4), 189–222.

Handley, K. and N. Limão (2017). Policy uncertainty, trade, and welfare: Theory and evidence for china and the united states. *American Economic Review 107*(9), 2731–83.

Havránek, T. (2015). Measuring intertemporal substitution: The importance of method choices and selective reporting. *Journal of the European Economic Association 13*(6), 1180–1204.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In

Handbook of international economics, Volume 4, pp. 131–195. Elsevier.

HOEKMAN, B. M. (2015). International cooperation on public procurement regulation.

Hoekman, B. M. (2017). International cooperation on public procurement regulation. In B. M. H. Aris Georgopoulos and P. C. Mavroidis (Eds.), *The Internationalization of Government Procurement Regulation*, Chapter 20, pp. 568–602. Oxford: Oxford University Press.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2016). The allocation of talent and us economic growth.

Huggett, M., G. Ventura, and A. Yaron (2011). Sources of lifetime inequality. *The American Economic Review 101*(7), 2923–2954.

Kalinova, B., A. Palerm, and S. Thomsen (2010). "oecd's fdi restrictiveness index: 2010 update". Technical report, OECD Working Papers on International Investment, 2010/03, OECD Publishing.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Koopman, R., Z. Wang, and S.-J. Wei (2014). Tracing value-added and double counting in gross exports. *American Economic Review 104*(2), 459–94.

Kutlina-Dimitrova, Z. and C. Lakatos (2016). Determinants of direct cross-border public procurement in eu member states. *Review of World Economics 152*, 501–28.

Larch, M., J. Wanner, Y. Yotov, and T. Zylkin (2017). The currency union effect: A ppml re-assessment with high-dimensional fixed effects.

Lindenlaub, I. (2017). Sorting multidimensional types: Theory and application. *The Review of Economic Studies 84*(2), 718–789.

Maggi, G. and A. Rodriguez-Clare (1998). 'the value of trade agreements in the presence of political pressures'. *Journal of Political Economy 106*(3), 574–601.

McDaniel, C. (2007). Average tax rates on consumption, investment, labor and capital in the OECD 1950-2003.

Mueller, F., K. O. Stahl, and F. Wachtler (2016). Upstream relationships in the automotive industry: A contractual perspective. Technical report, mimeo.

Onur, I., R. Ozcan, and B. Tas (2012). Public procurement auctions and competition in turkey. *Review of Industrial Organization 40*, 207–23.

Ramboll, M. C. and HTW Chur (2011). Cross-border procurement above eu threshold. Technical report, Ramboll, Copenhagen.

Rickard, S. J. and D. Y. Kono (2014). Think globally, buy locally: International agreements and government procurement. *The Review of International Organizations 9*(3), 333–352.

Sachs, D., A. Tsyvinski, and N. Werquin (2016). Nonlinear tax incidence and optimal taxation in general equilibrium.

Santos Silva, J. and S. Tenreyro (2006). The log of gravity. *The Review of Economics*

and statistics 88(4), 641–658.

Shingal, A. (2011). Services procurement under the wto's agreement on government procurement: Whither market access? *World Trade Review 10*(4), 527–49.

Shingal, A. (2015). Econometric analyses of home bias in government procurement. *Review of International Economics 23*(1), 188–219.

Shingal, A. and V. Ereshchenko (2018). Mapping government procurement provisions in preferential trade agreements:evidence from a new classification. mimeo.

Stantcheva, S. (2017). Optimal taxation and human capital policies over the life cycle. *Journal of Political Economy 125*(6), 1931–199.

Tas, B. K. O., K. Dawar, P. Holmes, and S. Togan (2017). Does the world trade organization government procurement agreement deliver what it promises?

Teulings, C. N. (2005). Comparative advantage, relative wages, and the accumulation of human capital. *Journal of Political Economy 113*(2), 425–461.

Timmer, M., B. Los, R. Stehrer, and G. de Vries (2016). An anatomy of the global trade slowdown based on the wiod 2016 release. Technical report, Groningen Growth and Development Centre, University of Groningen.

Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer, and G. J. De Vries (2015). An illustrated user guide to the world input–output database: the case of global automotive production. *Review of International Economics 23*(3), 575–605.

Ueno, A. (2013). Multilateralising regionalism on government procurement.

World Bank (2017). Benchmarking public procurement 2017.

Yamaguchi, S. (2012). Tasks and heterogeneous human capital. *Journal of Labor Economics 30*(1), 1–53.

Yotov, Y. V., R. Piermartini, J.-A. Monteiro, and M. Larch (2016). *An advanced guide to trade policy analysis: The structural gravity model*. World Trade Organization Geneva.